Causal Effects of Early Career Sorting on Labor and Marriage Market Choices: A Foundation for Gender Disparities and Norms *

Itzik Fadlon, *University of California, San Diego and NBER*

Frederik Plesner Lyngse, *University of Copenhagen and CEBI*

Torben Heien Nielsen, *University of Copenhagen and CEBI*

Abstract:

We study whether and how early labor market choices determine longer-run career versus family outcomes differentially for male and female professionals. We analyze the physician labor market by exploiting a randomized lottery that determines the sorting of Danish physicians into internships across local labor markets. Using administrative data spanning ten years after physicians' graduations, we find causal effects of early-career sorting on a range of life cycle outcomes that cascade from labor market choices, including human capital accumulation and occupational choice, to marriage market choices, including matching and fertility. The persistent effects are entirely concentrated among women, whereas men experience only temporary career disruptions. The evidence points to differential family-career tradeoffs and the mentorship employers provide as channels underlying this gender divergence. Our findings have implications for policies aimed at gender equality in outcomes, as they reveal how persistent gaps can arise even in institutionally gender-neutral settings with early-stage equality of opportunity.

* We thank Julie Cullen, Gordon Dahl, Claudia Goldin, Larry Katz, Henrik Kleven, Camille Landais, Petter Lundborg, Karthik Muralidharan, Claudia Olivetti, Linh Tô, Tom Vogl, Melanie Wasserman and seminar participants at UCSD, University of Copenhagen, Berlin Applied Micro Seminar, USC, McGill University/Boston College Labor Seminar, Harvard Medical School/Health Care Policy, Norwegian School of Economics FAIR Bergen, University of Bergen, University of Aarhus, BI Norwegian Business School Oslo, CEPR Public Economics, Essen Health Conference, IIPF Annual Meeting, the EEA Annual Meeting, and the NBER Summer Institute Gender in the Economy Meeting for helpful comments and discussions. We thank the Danish National Health Authority, De Regionale Råd for Lægers Videreuddannelse, and Dansk Telemedicin A/S for data access and helpful institutional knowledge. We also thank Carl Benjamin Simpson and Philip Nye for excellent research assistance. Fadlon: University of California, San Diego, Department of Economics, and NBER, 9500 Gilman Drive #0508, La Jolla, CA 92093-0508, fadlon@ucsd.edu; Lyngse: University of Copenhagen, Department of Economics, and Center for Economic Behavior and Inequality (CEBI), fpl@econ.ku.dk; Nielsen: University of Copenhagen, Department of Economics, and CEBI, thn@econ.ku.dk. We gratefully acknowledge funding from the Danish Independent Research Fund (grant no. 9061-00035B). The activities of CEBI are financed by the Danish National Research Foundation (grant no. DNRF134).

1. Introduction

Modern economies continue to struggle with gender disparities in the labor market. Gender gaps in performance and pay remain pervasive even among high-skilled professionals, such as lawyers, business and finance professionals, and physicians (Azmat and Ferrer 2017, Cullen and Perez-Truglia 2019, Sarsons 2019, Zeltzer 2020, Wasserman 2022). These gaps display a clear, common phenomenon: negligible disparities upon graduation, in line with their similar pre-market training across gender, pursued by a persistent widening pattern following the early-career stage and the subsequent family formation and arrival of children (Bertrand, Goldin, and Katz 2010, Goldin 2014, Blau and Kahn 2017, Juhn and McCue 2017, Kleven, Landais, and Søgaard 2019).

In this paper, we ask whether this persistent pattern among the highly-skilled represents a causal link from the early-career stage to long-run gender gaps in the labor market and spillovers to family choices. Specifically, in the context of the physician labor market, we establish whether and how early-career choices can determine longer-run labor market versus marriage market outcomes differentially for men and women. We study the sorting of Danish physicians into entry-level positions, which offers three main features.

First, placement into medical internships—i.e., physicians' first jobs—is governed in Denmark by a purely randomized lottery that provides a clean source of idiosyncratic variation in entry-level labor market sorting. This type of variation (that at the individual level) is required for identifying the causal effects of making differential choices during the early-career stage, which is a key input in a young professional's optimization problem. As we verify, students with the best lottery ranks, who are the ones that choose first, are effectively unrestricted in their choices and are assigned their highest priorities, whereas students with the worst lottery ranks, who are the ones that choose last and well after their choice sets have narrowed, are assigned their lowest priorities. We leverage this simple regularity to construct our control group (best lottery ranks) and our treatment group (worst lottery ranks), and we find that this generates large exogenous variation so that graduating physicians in the treatment group are much more likely to sort into internships in less desirable local labor markets and positions in the "first stage." We show that these positions offer inferior training and future career opportunities, such as lower-ranked educational programs, decreased affiliation with teaching hospitals, weaker professional networks, and higher likelihoods of locating in rural communities that display more traditional gender norms.

Second, we exploit a novel dataset that combines the formal lottery data we have digitized with a range of administrative datasets on all medical doctors in Denmark. These datasets cover information from medical registries on licenses and specializations and the Danish economic registers with information on location, employer-employee linkages, income flows from any reported source, education, and demographics. Importantly, we can link households using spousal and parent-child linkages to investigate family formation and fertility. Together, the data allow us to study a wide range of life cycle choices, in both the labor market and the marriage market, which provides us with the unique advantage of conducting a comprehensive analysis on the broad potential causal effects of early careers on work versus family tradeoffs. The data allow us to track our sample over a long period of up to ten years after the treatment.

Third, our setting readily lends itself to investigations of mechanisms in support of our main analysis. This is due to the information that maps the lottery ranks to choices, information on students' formal priority rankings over markets, and restricted data we obtained from the official government exit surveys in which interns assess their positions in different categories. The setting also conveniently creates differential variation along the various internship dimensions, specifically market location and ranked quality. These features allow us to shed light on the multi-dimensional treatment "bundle" that characterizes the early-career stage and to investigate potential sources of the long-run gender divergence we uncover.

Overall, we show that early-career labor market sorting has far-reaching causal effects on life cycle outcomes, from labor market choices of human capital accumulation and occupation to marriage market choices of matching and fertility. While men and women are subject to the same treatment, men experience only transitory career disruptions, and the persistent longer-run effects on all margins are entirely concentrated among women. The initial labor market sorting and the consequent choices in the decade that follows carry over to explain 10-14 percent of the projected earnings gap across male and female physicians three decades later. A key takeaway from our analysis is that persistent gender inequality still appears even in a context of a highly skilled merit-based profession with institutional early-stage equality of opportunity. The evidence underscores gender disparity in trading off career and family choices, as well as the mentorship employers offer in the pivotal early-career stage of young professionals, as channels driving the gender divergence in longer-run effects.

Our analysis is structured as follows. In its main part, we investigate the effects of initial career sorting on our two categories of longer-run outcomes: the labor market outcomes of human capital investment and occupational choice; and the marriage market choices of family formation and fertility.

We first find significant impacts on labor market choices. The advanced human capital investment that is most relevant in our setting is obtaining a medical PhD. This choice represents an occupational choice of a research career and provides, as we show, access to economically more favorable and prestigious positions, such as in university hospitals. While we find that males do not experience any adverse effects from the treatment, treated females are 25 percent less likely to make this investment. This impact alone can account for one-fifth of the observed gender-biased sorting into scientific careers, as compared to clinical positions, among physicians in our sample. Moreover, in terms of gendered occupational sorting, we find no effect on men but that treated women are more likely to sort into female-represented medical specialties. These specialties tend to have more temporal flexibility (Korreman 1994, Wasserman 2021), but it comes at the expense of lower financial returns, as we show.

To summarize the "very" long-run implications of these impacts, we use the surrogate index method (Athey et al. 2019) to project future earnings over the course of thirty years from graduation using our long panel of observational data. This allows us to address the regularity that the returns to major human capital investments could materialize only far in the future (e.g., 15 years after graduation in our medical PhD application, whereas our experimental data spans ten years). We show that the transitory variation in graduates' very first jobs alone explains 10-14 percent of the projected long-run gender gap in physician earnings in the third decade after graduation. Finally, we provide investigations that "unpack" the treatment bundle to assess drivers of these labor market effects. While there is evidence of some role for a position's ranked quality, the geography of the entry-level labor market can account for the bulk of the effects, which we show can be explained by a location's degree of rurality and affiliation with university hospitals.

We then investigate the interplay between the labor market and the marriage market by studying how the early-career stage can affect partnership and fertility choices, as this life stage represents formative years for family formation choices (Goldin and Katz 2008). We split individuals based on their partnership status at baseline since partnered and single graduates enter this stage with distinct family choice margins: partnered individuals already formed household units, whereas single individuals also match in the marriage market (which could be altered by the treatment).

We find significant impacts on fertility choices among the single graduates. With no effects on men, women in the treatment group exhibit an increase of 11.5 percent in their number of children, which is particularly driven by a higher fertility rate with an increase of 7.1 percentage points (pp) in the propensity to have more than one child (on a baseline of 45.2 pp). The absence of a similar impact on partnered graduates suggests this effect is less likely due to an underlying shift in household's family preferences, but instead relates to differential matching in the marriage market among single graduates. Indeed, we find effects on single women's matching patterns, where women in the treatment group end up in relationships with decreased assortative mating on age and education.

In the final part of the paper, we investigate mechanisms that can explain the gender divergence. First, we find strong evidence of gender differences in family versus career tradeoffs in response to the variation in early careers. We show that the women who have more children due to the treatment also invest less in human capital, and that their location decisions reflect family considerations as they show increased propensity to live near grandparents. This is consistent with women crowding out long-run career goals for more family-oriented choices as a result of unfavorable early-career placements. In comparison, men engage in career-oriented actions in response to unfavorable placements, which help them fend off potential adverse effects. These come in the form of search behavior in the labor market: men are more likely to travel further for work, thus displaying higher willingness to commute (consistent with Le Barbanchon, Rathelot, and Roulet 2021), and they are more likely to migrate their families across labor markets. As a second operative channel, we find evidence of an important role for the entry-level workplaces and their characteristics in initiating disparities. Whereas treatment group graduates of both genders similarly sort into employers with worse track records in placing their interns, women display a higher "sensitivity" to the employer characteristics in that they are more likely to end up in a less competitive subsequent job (for a given workplace track record). A specific workplace characteristic that emerges as important is the mentorship the workplace provides. We show that the treatment group is much less likely to be exposed to or assigned a female mentor, but that only treated female interns rank their mentorship experience and quality lower in response to the quasi-experiment. Lastly, the data are strongly inconsistent with differential preferences over entry-level positions as a channel. Males and females reveal very similar aggregate preferences in their choices over entry-level markets and positions.

Our advancement of the work on gender in modern high-skilled markets makes several contributions. Most directly, a key contribution of our analysis is to the long-standing work on gender inequality in economic outcomes and their underlying sources (see reviews and discussions in, e.g., Bertrand 2011, Goldin 2014, Olivetti and Petrongolo 2016, Blau and Kahn 2017, Lundberg and Stearns 2019).[1] We advance this literature by establishing the causality of early-career choices in initiating and perpetuating gender inequality and norms in long-run economic outcomes. We provide new evidence for how males and females differentially adapt to the tension between career and the family, which is most important in the study of gender (Goldin and Katz 2008, Goldin 2021), and show how early-career choices simultaneously cause disparities in both domains. We also offer insights into the mechanisms that could drive gender asymmetry in the effects of early careers. As our analysis reveals that significant gender inequality can emerge in a randomized lottery setup with embedded early-stage equality of opportunity, it has important policy implications. Policies for outcome-based gender equality cannot merely rely on leveling the starting playing field, but they should also target the way in which opportunities and choices evolve over the formative stage of the early career. Our analysis of mechanisms offers some initial guidance in that direction.

We additionally contribute to the classic labor economics research that has highlighted the potential importance of early-career stages in determining long-run life cycle trajectories. This work has considered the role of search and job mobility, human capital investments, as well as on-the-job learning and skill

[1] Recent important studies in this active research on underlying channels investigate the role of job search and labor market preferences, social interactions, personality characteristics, and family obligations. These include, among others, Gneezy et al. (2003), Niederle and Vesterlund (2007), Bertrand et al. (2010), Buser et al. (2014), Azmat et al. (2016), Card et al. (2016), Field et al. (2016), Azmat and Ferrer (2017), Bursztyn et al. (2017), Caliendo et al. (2017), Buser and Yuan (2019), Cai et al. (2019), Cullen and Perez-Truglia (2019), Exley and Kessler (2019), Iriberri and Rey-Biel (2019), Kleven et al. (2019a), Kleven et al. (2019b), Cheng (2020), Porter and Serra (2020), Ginther et al. (2020), Le Barbanchon et al. (2021), Cortés et al. (2022).

accumulation (see, e.g., Topel and Ward 1992, and reviews in Weiss 1986 and Rubinstein and Weiss 2006). We contribute to this broad line of research by providing a novel, purely randomized source of idiosyncratic variation for identifying the causal effects of early-career choices.[2] This type of variation can be useful in other important economic questions. For example, with a focus on market design, Arora et al. (2021) concurrently study how shifting the Norwegian system of medical internship allocation from lottery-based to market-based has impacted the quality of employer-employee equilibrium matching using a metric based on earnings five years out. In our Danish setting and analysis, we provide causal evidence on the far-reaching, longer-run impacts of early-career choices on a wide range of economic outcomes. We show how the impacts of initial career sorting cascade from labor market and human capital choices, to marriage market and fertility choices, to the important tradeoffs across them.

Finally, we speak to the mounting recent evidence that highlights geographic location in determining life cycle outcomes, from education, to economic well-being, to health.[3] We contribute to this strand of the literature first by finding a causal determinant of the household's choice of geographic location in the long run—namely, early-career labor market sorting (as we find strong lingering effects on location). This choice directly affects the local labor market in which the household operates and the amenities available to the family. Second, our findings identify a pathway by which the location in which individuals operate can shape behavior and welfare. We show that the treatment effects we identify can be attributed to location effects, thereby highlighting the potential causal role of location via geographic sorting in the early career, which, in turn, affects households' long-run opportunities and economic life trajectories.

The rest of the paper proceeds as follows. Section 2 describes the institutional setting. Section 3 describes the data sources we use and baseline patterns in internship choices. In section 4, we set out our empirical framework. Section 5 describes the nature of the quasi-experimental treatment in terms of the first stage. Section 6 provides the evidence on the longer-run causal effects of early-career choices and their gender divergence. Section 7 investigates mechanisms underlying this divergence. Section 8 concludes.

2. Institutional Setting

In this section, we provide a detailed description of the context of our analysis. We describe the course of postgraduate professional experience and training of Danish physicians, which captures the early

---

[2] Related but distinct work had studied aggregate variation, in terms of entering the labor market in a recession (see von Wachter 2020 for a review), in contrast to the idiosyncratic variation that we study here. The former identifies the effects of changes to the choice set that come from bad economic times. In comparison, the latter, with variation at the individual level, identifies the causal effects of making different choices within a given distribution of options, i.e., a given choice set. As such, these effects form a key input in an individual's optimization problem of early-career choices. In that sense, our analysis resembles the economics of education literature that uses idiosyncratic exogenous variation (e.g., based on grade cutoffs) to analyze the returns to different choices of field of study (e.g., Kirkeboen et al. 2016).

[3] See, for example, Chetty and Hendren (2018) and Finkelstein et al. (2016) for the U.S., and Damm and Dustmann (2014), Laird and Nielsen (2016), and Eckert et al. (2019) for our context of Denmark.

stages of their careers, and we elaborate on the process of matching to medical internships in Denmark, which provides the grounds for our causal analysis.

2.1. Physician Training: Broad Overview

The timeline for Danish physicians' training process is generally typical of other OECD countries.[4] Following medical school, graduating physicians begin the period of their *residency*. The residency represents a period of on-the-job training during which physicians make pivotal human capital investments and occupational choices (such as medical specialties) that determine their career paths. The various stages of the residency period (illustrated in Appendix A) are as follows.

The initial stage of the residency is the *internship*, which typically lasts one to one and a half years. The internship represents the entry-level labor market for physicians. It stands as physicians' first effective medical experience, and it determines their initial exposure to practical knowledge and career opportunities. The key institutional feature, which we exploit as the basis for our identification, is that a random lottery underlies the placement to internships. We provide more contextual information on the internship and a detailed description of the assignment process in the next subsection.

After they complete their internship, the starting physicians are allowed to practice medicine independently, that is, without the supervision of a senior physician. At that stage, the physicians engage in a process of job search as well as human capital investments (specifically, pursuing a medical PhD) that will govern their later positions and career paths. All positions after the internship period are matched in a standard competitive labor market.

In the immediate stage after the internship, the physicians apply for different *introductory positions*, which typically last one year each. The physicians must complete at least one such position within their future specialty of interest. This would then qualify them to apply for a *main position* within a specific choice of medical specialty. The main position represents the last stage of the residency, whereby the choice of specialty is typically an absorbing state in relation to the physicians' future careers.

Main positions can be highly competitive, and hence physicians' success in this stage is influenced by their choices of investment and training up to that point. Specifically, practical experience from relevant introductory positions and further academic education by obtaining a medical PhD degree are key determinants. In the longer run, a PhD degree could further qualify a physician for a broader set of higher-paying competitive positions (as we show later in the data), such as positions at university hospitals and prestigious positions of chief specialists. Upon the completion of the residency, physicians receive their specialty license and continue on to their future careers.

[4] For the institutional structure in EU countries, for example, see EU Council Directive 75/363/EEC.

2.2. Internship: Source of Identifying Variation

The internship following medical school provides our source of exogenous variation in initial job market sorting of physicians. The graduate is assigned to an internship position by being matched with a hospital department—that in practice represents a workplace—which is responsible for facilitating an educational program for the intern. The interns are supervised by formally assigned mentors, including a senior supervisor and the head of the educational program at the specific workplace.

Training Content. The internship positions aim to provide hands-on work experience and have the physicians accumulate practical knowledge and skills through learning by doing. That is, the purpose of the internship is to bring the theoretical knowledge from medical school into clinical practice by having the intern integrated into the daily work routines of a given hospital department. Legislation defines learning goals for the educational programs under several categories of virtues a physician should possess: medical expertise, communication, health promotion, collaboration, ethics, leadership, and academic merit (National Board of Health 2009). The medical expertise portion is meant to assure that an intern engages in all aspects of medical care, including diagnostics, examinations, implementation of procedures, treatment protocols, medical complications, resuscitation, and treatment of acute and chronic patients. Interns gain this experience by treating patients, interacting with patient relatives, and working with the multiple healthcare professionals who provide patient care. For the interns to complete the program successfully, their supervisors must sign off that they meet the expected standards of all the learning goals. By the end of the internship, interns evaluate the program and their experience via formal external exit surveys.

In terms of its structure, the internship consists of bundles of half-year primary positions at hospitals followed by secondary positions at primary-care practices. By definition, internships are tied to geographic regions and their hospitals. Institutionally, the healthcare system in Denmark is organized such that Danish counties (with a total of 16) represent the local healthcare market (which bears similarities to Hospital Referral Regions in the U.S.). We note that spatial variation in entry jobs for physicians is typical of postgraduate medical training positions in other developed countries, such as the U.S., and is a main dimension by which the training programs are categorized (see, e.g., Brotherton and Etzel 2018).

Internship Assignment. Internship positions are periodically created by the Danish National Health Authority (NHA), with respect to national demand for healthcare professionals, to accommodate all graduating students. Specifically, prior to every graduation round (twice a year), the NHA requires medical schools to report how many students will graduate in that round. The NHA then guarantees to create a number of internship positions of at least that amount. Finally, the positions are designed to distribute proportionally across the local labor markets (i.e., counties) based on their population share.

The key institutional feature we exploit for identification is that a *randomized lottery* governs the placement into internships. For every graduating cohort, a public notary performs a lottery that allocates a random number to each graduating student, which sets the ordering of the matching process for that cohort. We capture a graduating physician's relative position in the matching order by mapping a lottery number to its rank relative to the lottery numbers of the graduate's cohort. We refer to it as the "lottery rank."[5]

The exact implementation of assignment to internships based on the lottery has changed over the years, but it has been continuously designed so that a better lottery number (of a lower rank) guarantees a student a more favorable position in the allocation process. We leverage this simple yet powerful feature and pool all graduating cohorts to maximize power. We show in the next section (Section 3) that the patterns of allocation of students to internships remains very similar over time. To give context, prior to 2008, the NHA first allocated students to counties based on the order of their lottery numbers in the primary stage of the placement process. Having been assigned their lottery numbers, the graduating students then compiled their list of priority over all the Danish counties. Next, they were matched with their highest-ranked county among the counties with remaining open positions when their time in line to make a choice arrived. Later, each county matched its assigned students with the internship positions (across the county's hospitals) that were created in that round, based on student choices in the order of their initial lottery number. In 2008, when the system was digitized, the process simplified into a single combined step, where interns make a county-hospital choice in the order of their lottery number from the positions available nationally (known as random serial dictatorship, see Abdulkadiroğlu and Sönmez 1998).

3. Data and Patterns of Internship Choices

3.1. Data

We combine several administrative datasets that contain information on all medical doctors in Denmark and their households. We use the *Educational Registers* starting in 1980 to identify all students ever enrolled in a Danish medical school through 2017. Our analysis population for the lottery quasi-experiment is identified using information starting from 2001 on the internship lotteries, which we obtained from the physical archives at the Danish National Health Authority and digitized. We link these records using person-level identifiers with the following register datasets on the data servers at Statistics Denmark.

The *Danish Authorization Register* provides us with information through 2017 on registrations of medical licenses and specializations, which capture occupational choice in our setting. The economic registers (up to 2019) include administrative information on geographic location (up to 2019), employers and employer-employee linkages (up to 2017), income flows from any reported source, including wage earnings, self-employment income, government benefits, and capital income (up to 2017), demographics, including age and gender (up to 2019), and education, including information on degrees achieved and high

[5] This normalization permits a comparison between individuals with bad lottery numbers and individuals with good lottery numbers across cohorts of different sizes. Appendix Table E.2 provides estimates that include graduation round fixed effects for robustness.

school GPA (up to 2017). We are able to link households using spousal and parent-child linkages (up to 2019) to study matching in the marriage market and fertility choices, which is important for our purposes.

In addition, we obtained confidential information from the internship exit surveys, which are processed at restricted research servers at the University of Copenhagen. With permission from the official governmental body, the Regional Councils for Physicians' Postgraduate Education (De Regionale Råd for Lægers Videreuddannelse), these data were obtained from a private IT company, Dansk Telemedicin A/S, which administers the data on all postgraduate educational positions for physicians in Denmark. From 2008, due to the digitization of the internship selection process, we are able to link the lottery numbers to the exit surveys and to the exact positions at the specific hospital departments. In these externally conducted formal government surveys, the interning physicians assess their workplace in a series of questions that are clustered into topic-based categories. Besides overall assessment as a category, we will make particular use of the category related to the workplace mentorship. We report the full list of survey questions in Appendix G. These surveys also include details on the interns' supervisors and educational program chairs, which we use to identify the gender of the interns' mentors.

3.2. Patterns of Internship Choices

Location Choices during Internship. The effective choice and assignment patterns of the medical internships display close similarities over the years, which is consistent with students' reluctance to intern in remote and rural areas. The motivation underlying the randomization-based placement process (see Danish Ministry of Health 1989) had been to solve the joint problem of excess supply of young physicians around the university hubs as well as physician shortages in rural areas (which is a broader concern and a common policy target across OECD countries; see, OECD 2012, Ono et al. 2014). Throughout the years, geographic dispersion and relocation of graduating students have been a key dimension of variation that the lottery has created across the lottery rank distribution. To illustrate this, we calculate, for each student, the distance between their municipality of residence at the time of the lottery and their municipality of work at the time of the internship, which captures their "relocation distance." To put it in context, we note that graduating students reside near the major university cities in which medical schools are located in Denmark (Aarhus, Copenhagen, and Odense). Hence, short relocation distances imply staying in the vicinity of the urban labor market where the student was educated, and long relocation distances typically imply placement in internships that are located in rural areas.

Panel A of Figure 1 plots a graduating student's relocation distance against the student's lottery rank, where we split cohorts around 2008 (when the exact allocation process changed due to digitization). There is a clear gradient such that the relocation distance for those with better lottery numbers (lower ranks) is significantly shorter than for those with worse lottery numbers (higher ranks). This mirrors the underlying motivation for the lottery-based system, as it reveals interns' distaste for locating in rural labor markets when they get to choose.[6]

Strategic Choice Considerations. It is useful to discuss some potential choice and prioritization considerations, which could result from the incentives embedded in the choice processes we described, and to investigate how they play out in practice. That said, as will become clear in our research design, it is important to note that strategic behavior is not going to affect the validity of our identification of the effects of initial labor market sorting. This is because our choice for the main research design rests on reduced form effects of the randomized lottery numbers. Still, describing these aspects is potentially informative for understanding the empirical context and for interpreting our findings. We use the information on the full formal rankings of local labor markets provided by the earlier cohorts in these investigations.

Given the structure of the matching process, individuals' equilibrium best-response strategy at each stage is to choose the option that maximizes their expected utility payoff based on their individual preferences and their expectations of other students' equilibrium play. For the later cohorts, this simply implies choosing their most preferred option among the options that are still available at the time they make their choice. For the earlier cohorts, there are additional potential considerations to take into account. To the extent that differential job aspects within a county play a role in ranking preferences (that is, aspects that go beyond the local labor market and its typical internship-related characteristics), the process implies that, at the first step of ranking counties, some consideration may be given to one's place in line for making a choice. For example, it may be preferable (along some job dimension) to be first in line in a worse labor market than last in line in a better labor market.

To test how this conjecture may play out in practice, we consider the rankings by those with the best lottery numbers as compared to the rankings by those with the worst lottery numbers. Specifically, we use graduates with lottery ranks in the highest 30 percent and the lowest 30 percent to be consistent with our main analysis that follows. To the degree that students view their position in line for making a choice within a market as important, we would expect systematic differences in rankings over labor markets across the two groups. If, on the other hand, the choice of local labor market is what dominates students' preferences regarding where to intern, we would expect similarities in their overall rankings. Panel B of Appendix Figure C.1 compares the average rankings of labor markets across the two groups. Each dot represents a local market, and we plot the fitted line as well as the 45-degree line, which is the benchmark under non-differential rankings. We also report the slope of the fitted line, where the benchmark null of non-differential rankings is one. The figure is consistent with the second hypothesis, i.e., that the choice of

[6] The persistence of location preferences over the years, as they are revealed through choices, can be also shown in the following way. Let us characterize the desirability of a labor market (i.e., a county) based on the average lottery rank of the interns who choose to sort into it. We construct these desirability rankings for both earlier and later cohorts and compare across them in panel A of Appendix Figure C.1. Locations are effectively valued over the years to a similar extent.

labor market itself leads students' rankings in the first step of the allocation process. The average rankings of markets across the two groups line up around the 45-degree line, and we cannot reject the benchmark null of a coefficient of one. The importance of location in students' preferences and choices is further underscored later when we analyze the quasi-experiment's first stage across lottery rank groups.[7]

4. Empirical Framework

4.1. Verification of Lottery

As the basis for our empirical analysis, we establish the validity of the lottery in terms of random assignment. In Appendix Table B.1, we run specifications that regress the graduating physicians' lottery rank on baseline characteristics available in our data. These include gender, age, an indicator for having a registered partner, number of children in the household, and high school GPA rank. Consistent with random assignment, we find that these regressions have no predictive power. This is the case whether we test the significance of the coefficients individually or jointly. In the appendix table, we also run the corresponding specifications separately for males and females, with similar conclusions. This sets the grounds for our research design that we turn to next.

4.2. Nature of Assignment to Entry-Level Jobs

To motivate the choice of our research design, we first describe the practical nature of the internship assignment. To do so, we use the information on the binding pre-placement rankings of all local labor markets solicited among the earlier cohorts as part of the allocation process, and we study the mapping between choice rankings and placements as a function of the lottery. Specifically, we plot individuals' pre-placement ranking of the local labor market they were assigned to in practice (where 1 is highest priority) against the percentile rank of their lottery number within their graduating cohort.

Panel B of Figure 1 shows a few key patterns in this relationship. First, as expected by design, there is a clear gradient such that graduates with higher lottery ranks (worse numbers) are assigned their lower-ranked priorities. More interesting, however, is the market-clearing pattern of the available slots against graduates' preferences. We see that, in equilibrium, there is a clear non-linearity: there is a virtually flat region in the vicinity of the best lottery ranks and a steep slope in the vicinity of the worst lottery ranks. By the nature of the assignment process, students with the best lottery ranks are effectively unrestricted in their choices. As they are the ones who make the choices first, their highest priority options are still available when they make a choice, and therefore they end up being assigned their first priority. Then, as the lottery

[7] In addition, local labor markets and the typical characteristics of the jobs they offer have aspects that people may agree upon ("vertical" quality, e.g., interning in a teaching hospital) and aspects that could be individual specific ("horizontal" quality whose valuation can differ across individuals, e.g., a county's proximity to family). We investigate the degree to which the labor markets rankings are agreed upon among graduates in Appendix C.1 by comparing the average rankings of labor markets across a random split of our analysis sample. We find them to be very similar, implying a high degree of agreement among randomly drawn graduates.

rank increases (that is, draws worsen), the set of available choices increasingly narrows. As a result, those with the worst lottery numbers are most restricted in their early-career choices, and therefore they end up making choices that are lowest on their priority list. These patterns guide our choice of research design.

4.3. Research Design

To analyze how early-career sorting affects longer-run life cycle outcomes, we employ a straightforward design based on the randomized lottery where we compare outcomes of a treatment group to outcomes of a control group. As natural experimental groups, we define the "control" group to be individuals with the best lottery ranks (below a certain lower cutoff rank), as we have seen they are essentially unaffected by the lottery; and we define the "treatment" group to be individuals with the worst lottery ranks (above a certain upper cutoff rank), as we have seen they are the most affected by the lottery.

Our choice of research design, which compares outcomes of a treatment group to outcomes of a control group, provides a standard and intuitive empirical framework, with treatment effect coefficients that are economically directly interpretable. In addition, it maximizes the differential treatment intensity across the differentially affected experimental groups since it compares individuals who are most restricted to those who are least restricted in their choices. Finally, it does not impose functional form assumptions on the underlying relationship between outcomes and lottery ranks (specifically, it does not use the common linear in rank specification where linearity seems less appropriate given the patterns in Figure 1). Still, we also run the corresponding specifications that are linear in lottery rank (in Appendix Table E.1).

In constructing our experimental groups, we need to make a choice of upper and lower lottery rank thresholds, which we do in the following way. First, to keep the experimental groups balanced with similar size, we use symmetric thresholds from above and below. Second, we pivot the analysis around the 30 percent most treated and least treated, i.e., with cutoff ranks 0.30 from below and 0.70 from above (as illustrated by the vertical lines in panels A-B of Figure 1), and we vary this bandwidth from 20 to 40 percent in Appendix Table E.1. This choice trades off increased power from higher treatment intensity with decreased power from reducing sample sizes, which is the reason we investigate a broad range of 20 pp in lottery ranks for the robustness analysis of bandwidth choice in Appendix Table E.1. For completeness, the appendix table also reports the effects on the "middle" group of graduates with lottery ranks in the intermediate range.

We note that while we discuss our main results as the comparison between the treatment group and the control group, we also leverage comparisons to the middle group. As we will show below, the middle group and the treatment group are differentially affected by the treatment, in terms of the first stage, on different dimensions of the treatment, which is naturally multi-dimensional. This will give us the opportunity to unpack the treatment bundle.

Estimating Equation. With this design, we identify the causal effects of the internship lottery using the following estimating equation:

$$y_{it} = \sum_{\tau} I_{\tau} \times \alpha_{\tau} + \sum_{\tau} I_{\tau} \times Treat_{i} \times \beta_{\tau} + \varepsilon_{it}. \tag{1}$$

In this specification, y_{it} is the outcome of interest for individual i at time t; τ is the year relative to the lottery; and I_{τ} is a vector of indicators of time relative to the lottery in years. The variable $Treat_i$ is an indicator for being in the treatment group or in the control group. In our main analysis, we focus on the latter half of our data horizon (years 6-10) to focus on longer-run outcomes and since some key choices emerge several years after graduation (such as advanced education completion, which starts materializing after approximately five years). We cluster standard errors at the individual level.

Our parameters of interest are the elements of β_{τ}, which estimate the causal effects of the lottery treatment up to ten years. We summarize the average longer-run effects over years 6-10 using the following standard estimating equation:

$$y_{it} = \alpha + \beta \times Treat_{i} + \varepsilon_{it}, \tag{2}$$

where β captures the average longer run treatment effect.

Analysis Sample. Appendix Table B.2 describes our analysis sample and provides summary statistics for our treatment and control groups. Overall, the two groups together consist of 6,076 physicians. Some particular characteristics that would prove useful for later discussions include the average age of our subjects at the time of the lottery, which is about 28.5, and that about half of our subject pool have a partner at the baseline period. Approximately 60 percent are female, with 2,396 males and 3,680 females in our sample. Summary statistics that split the sample by gender are also provided in Appendix Table B.2.

5. The First Stage: Internship Period Exposure

How do the lottery ranks translate to differential exposure to internship characteristics? As a starting point, we characterize the nature of the treatment by investigating the effects of the lottery on the entry-level labor market positions doctors sort into. This serves as the first stage analysis, which sets the basis for interpreting the long-run effects of the lottery. As in any natural experiment, it is important to note that this treatment is a "bundle." It includes aggregate characteristics of the local labor market to which interns are allocated and characteristics of the specific internships they are matched with. We now turn to describing these characteristics, and then we discuss how the setting lends itself to exploring the multi-dimensionality of the treatment.

5.1. Tradeoff: Position Local Labor Market and Program Ranked Quality

There are at least two key underlying dimensions that pertain to the internship assignment. The first dimension is location, which, as we discussed in the institutional background, comes from the motivation of the lottery-based policy to counteract students' reluctance to intern in rural areas. We again use our measure of relocation distance, which, in our context, maps to the likelihood of being placed in a less desirable rural labor market. The second dimension is a ranked quality of the educational program itself (the specific workplace/position), as reported by interns in the exit surveys. We use the ranking of a position's overall assessment, whereby graduates are asked about their overall evaluation of the educational experience in terms of the program's effort, quality of training, and their own professional development (see Appendix G). In the current analysis, we use later cohorts from after the digitization of the system for whom we have detailed information available on both dimensions of the internship allocation. We note that the value of the ranked quality is that it can capture, via the reported experiences of interns, measures that are not directly observed. Indeed, recent studies (e.g., Alsan et al. 2019, DellaVigna et al. 2020) leverage surveys to offer insights into channels like we do here, as the surveys conducted by the NHA neatly cover such dimensions. Still, it is useful to corroborate these assessments against external measures. We have data from external inspections that the NHA conducts to assess the quality of the educational programs. These were performed by independent inspectors and cover a majority subset of programs. Appendix G.2 shows that these external ranking are highly predictive of exit-survey rankings.

Panels C-E of Figure 1 investigate the relationships of the two measures with the lottery. Panels C and D first separately plot averages of these measures as a function of lottery ranks for each of fifty equal-sized bins. Panel C measures the relocation distance in kilometers from the municipality of residence (at the time of lottery draw) to the municipality of the internship workplace. Panel D measures ranked quality at the hospital department level, i.e., the "workplace." We use the leave-one-out mean (to avoid mechanical correlations with own lottery rank) of the overall evaluation of graduates who interned in a given workplace, which we normalize by the mean and standard deviation to create a z-score. Panel E then aggregates this information across our "control" group, "treatment" group, and "middle" group, and it plots the averages of the two dimensions simultaneously for each of our experimental groups. This figure bears similarities to an "offer curve" if the internship bundle is to be thought of as a consumption bundle: the curve maps individuals' choice of a multi-dimensional bundle for an increasingly narrow choice set. Corresponding figures for a finer set of groups of lottery ranks are provided in Appendix D.

A clear tradeoff pattern arises in panel E of Figure 1. First, we see that, as expected, the "control" group, for whom there are virtually no restrictions, chooses internships that are closest to their medical school's urban hub *and* in positions that are higher ranked. The "treatment" group who is most restricted suffers on both margins, as they end up choosing remaining positions that are *both* in remote geographic locations and are of lower-ranked quality. Finally, the "middle" group's choice contains important revealed preference information. On average, the "middle" group is on par with the "control" group in terms of distance, whereas the "middle" group is on par with the "control" group in terms of a position's ranked quality. This (along with the finer version in Appendix Figure D.1) illustrates a lexicographic nature of preferences for internship locations among graduates: they are willing to choose the lowest-ranked positions within a local market for the opportunity to intern in a more desirable geographic location.

A useful feature of these patterns is our ability to shed light on different dimensions of the treatment. Intuitively, in studying the longer-run causal effects, a comparison between the middle group and the control group would assess the role of a position's educational ranked quality (for a given set of geographic markets), and a comparison between the treatment group and the middle group would assess the role of geographic markets (for a given level of a position's educational ranked quality). We formalize this simple intuition in Appendix D.1. Under standard assumptions of additivity (i.e., lack of complementarities across the two dimensions) and exclusion (i.e., that the two dimensions capture the bulk of the variation relevant for long-run outcomes), this analysis offers a complete decomposition of the total effect, whereas the decomposition would be partial if these identifying assumptions were meaningfully violated.

Finally, panel F of Figure 1 shows that the tradeoffs are similar across gender since the gender-specific offer curves reveal a similar shape. This implies that potential differences in longer-run effects across gender could not be attributed back to differential first stages or how they may have differentially translated on average to the treatment intensities across dimensions.

5.2. What Characterizes Local Geographic Markets?

If geographic markets turn out to matter, it is important to provide a description of their key characteristics in the context of our quasi-experiment. To proceed, we construct a measure of a geographic labor market's desirability. We characterize the desirability of a labor market, i.e., a county, based on the average lottery rank of the interns who choose to sort into it. This captures the aggregate regularities that a market is revealed as more desirable if it is chosen by individuals with better lottery ranks, and a market is revealed as less desirable if it is chosen by individuals with worse lottery ranks. We then use these rankings to partition the markets into two groups: more desirable and less desirable local labor markets.[8] Finally, we study correlations of market desirability with characteristics that could capture aspects of its added value in terms of, e.g., on-the-job training and future career opportunities through professional networks.

A key such characteristic is the extent of attachment to university (or teaching) hospitals. Notably, we find, in panel A of Table 1 (column 1), that being assigned to a less desirable local labor market is associated with 31 pp lower likelihood of interning in a university hospital (on a baseline of 40 pp). Leading university hospitals, which are typically located in local labor markets in the vicinity of larger urban areas, are well known to be the institutions where skill-intensive and highly specialized procedures are performed, state-of-the-art technologies are first adopted, and innovative medical research is conducted. By definition,

[8] This market partition is similar if we split locations based on the average pre-placement rankings of local labor markets using the information on students' solicited priority lists among the earlier cohorts.

university hospitals aim to provide the highest quality on-the-job training to new physicians. Moreover, since key players in the medical field often work and mentor in these hospitals, interning in a university hospital could provide more favorable exposure to professional networks.

We use the administrative patient register data to illustrate these points. Panel B of Table 1 compares university hospitals to non-university hospitals on different dimensions. First, in terms of scale and the level of technology, we find that university hospitals offer exposure to more patients and types of procedures as well as to more advanced medical technologies, based on common measures in the literature such as the prevalence of MRI scanners (see, e.g., Bhattacharya et al. 2013). As a measure of high-quality training and favorable professional networks, we consider the share of high-seniority colleagues. For each hospital, we look at the share of physicians that already obtained their specialty who hold a medical PhD (out of all physicians that had obtained their specialty). The logic behind this measure is that physicians who hold a medical PhD tend to occupy the key positions in the field, and we find that university hospitals rank higher on this dimension as well. This difference can also capture variation in the type of role models young physicians are exposed to and mentored by in the internship. Lastly, university hospitals offer exposure to wider specialized knowledge through the presence of a wider range of medical specialties.

Another important dimension of a local healthcare market, which is again related to the motivation underlying the lottery-based allocation to counteract distaste for rural locations, is the degree to which the lottery affects the probability a graduate would intern in a rural community versus an urban community. We follow the formal definitions used by the Danish Economic Councils (2015) that are based on classifications at the level of municipalities (which are sub-divisions of counties). The urban-rural divide is frequently used in the discussion of localities more broadly and in the characterization of healthcare markets and physicians' postgraduate training more specifically.[9] Panel A of Table 1 (column 2) shows that a locality is 61.5 pp more likely to be rural if it is located within a less desirable local labor market.

Panel C of Table 1 provides a characterization of rural municipalities on several dimensions that relate to demographics, amenities, and features of the healthcare market in which the graduates intern. Rural areas are characterized by populations that are less educated, sicker, and more reliant on welfare. These municipalities have worse economic conditions and amenities (in terms of income, home prices, tax revenues, and local recreational expenditure). Finally, with our focus on gender, we look at measures that could capture gender-related norms. In terms of traditional household roles, we find that, in rural areas, females are much more likely to take parental leave, with the opposite pattern for males. In terms of local

[9] For example, this characterization of healthcare markets in the U.S. is structurally embedded in the operation of Medicare and its pricing schemes (see, e.g., Sloan and Edmunds 2012). Additionally, geographic imbalances in the form of physician degree of concentration in rural versus urban areas are a pervasive phenomenon across the developed world, and countries have taken several policy measures that aim to address physician shortages in rural areas (see Simoens and Hurst 2006, OECD 2012, Ono et al. 2014). One example is Medicare's Health Professional Shortage Area (HPSA) Physician Bonus Program in the United States.

representation, we find that the share of elected officials who are female is lower in rural areas. These are consistent with general priors as they suggest that rural areas may be more gender-stereotypical overall.

We have highlighted two features of the geographic market composite that seem to us to stand out—affiliation with university hospitals and degree of rurality—which we have shown to be strong predictors of some entry-level job dimensions that are important in the early-career stage. Of course, other features beyond these two could be (and likely are) a part of the local labor market composite. We later assess the degree to which our characterization of a location's desirability based on these two features is comprehensive or, in other words, the degree to which we have "pinned down" the first stage. To do so, we will investigate the extent to which variation in them can explain the longer-run treatment effects on labor market outcomes.

5.3. A Bridge from the First Stage to the Longer Run

As a segue from the first-stage treatment to the longer-run impacts, we consider the dynamics in a household's geographic sorting, given that the internship allocation system is strongly governed by location. The importance of individuals' choice of geographic location stems from the fact that it can directly affect both the local labor market and the marriage market in which the individual operates. Figure 2 illustrates the dynamic effects of the lottery on the probability of sorting into differentially desirable local labor markets throughout our entire analysis window. It plots the β_τ estimates from equation (1) for years 0 to 10, along with their 95-percent confidence intervals, where the x-axis denotes the year relative to the lottery.

The early years mechanically capture the first stage effect on the internship placement, particularly year 1, which is the period where the internship placement is in full effect (given the timing of the internship relative to the lottery and the end-of-year timing of data reporting in the registers). We see that receiving the worst lottery ranks leads to a large 18.4 pp increase in the probability of interning in less desirable healthcare labor markets (on a counterfactual of 11.6 pp). Notably, focusing on the longer run, the figure reveals that the lottery has important lingering effects that persist throughout the years. Ten years after the lottery—long after the internship itself—individuals in the treatment group are 6.5 pp more likely to sort into less desirable local labor markets relative to a counterfactual of 16 pp.

We then split the sample by gender. Panel B first shows that both males and females have similar sorting patterns at the internship period; that is, they have a similar first stage. This again means that differences across gender in long-run effects cannot be traced back to potential differential assignment in the treatment stage. Then, studying the dynamics of geographic sorting, an important asymmetry unfolds: the long-run effect is entirely attributed to women. While men do not display effects in the long run, women display a 9.8 pp increase in the propensity to sort into less desirable local labor markets on a baseline of 14 pp. This gender divergence is a precursor to our main analysis that follows.

6. Long-Run Effects on Life Cycle Choices by Gender

With this setup, we now turn to our main analysis and investigate how the internship placement affects life cycle choices up to ten years after graduation differentially for males and females. We divide the longer-run analysis into two categories of household decisions: (i) labor market choices: human capital investment and occupational choice, and (ii) marriage market choices: household formation and fertility.

6.1. Labor Market Choices: Human Capital Investment and Occupational Choice

Human Capital Investment. We first study a classic human capital investment of obtaining a medical PhD, which also represents an occupational choice of a research track and scientific career in our setting. This human capital choice stands as an important upward career move, as it provides access to economically more favorable and prestigious positions (Korremann 1994), e.g., in university hospitals. Using the population-level register data, panel A of Figure 3 shows an investment pattern of a classic labor economics shape: it illustrates the association of obtaining a PhD with early lifetime investments, in terms of foregone income, and with high returns later in the life cycle. We note that this is a lengthy investment in that its net returns manifest only late, a point that will become relevant later.

Panel A of Table 2 studies the likelihood of obtaining a medical PhD, and thus also the corresponding sorting into a scientific career track. It provides estimates for β_τ using equation (1), starting from year 6, which is when PhD completion begins to materialize following graduation from medical school. The results reveal a clear gender divergence. Males do not experience any adverse effects as a result of the treatment. However, females in the treatment group have significantly lower propensity to make this human capital investment. By the end of our analysis period, females' lower investment rate amounts to a large decline of 5.4 pp in obtaining a PhD on a counterfactual of 21.3 pp.

Our findings directly relate to gender-biased sorting into scientific careers, with gender inequality in science being a well-known phenomenon of concern in the developed world (Holman et al. 2018, Huang et al. 2020). We calculate among our subject pool that the male-female gap in holding a medical PhD ten years after graduation is 8.24 pp. This implies that the treatment effect can account for 20 percent of the observed gap.[10] These large effects are attributed to variation in the short internship period alone (out of the lengthy process to become a physician), underscoring just how important choices during the very early career stage could be over the course of the life cycle.

As key positions in the medical field are attached to university hospitals and tend to be held by medical PhDs, a related result pertains to physicians' affiliation with university hospitals in the long run as a function of our quasi-experimental variation. Panel B of Figure 3 first illustrates how affiliation with

[10] These assessments assume (as we show later in this subsection) that it is individuals with the worst lottery numbers, i.e., those included in our treatment group, who are adversely affected on this margin. As they compose 30 percent of the sample, our calculation is performed as follows: $(5.42 \times 0.3)/8.24 = 0.20$.

university hospitals is a strong indicator for economically favorable career trajectories in the long run. Panel B of Table 2 then presents the results, showing that, consistent with our findings so far, there are no effects on males, but there are meaningful adverse effects on females in the treatment group. In the longer run, they are 6.5 pp less likely to be affiliated with a university hospital on a counterfactual of 40.5 pp.

Gender-Represented Specialties. We further investigate the differential occupational choice that could reinforce gender norms by studying sorting into gender-represented occupations. We classify medical specialties—which represent "occupations" in our setting—based on the share of females within a specialty relative to their overall proportion. "Female-represented specialties" are defined as specialties with a female share that is higher than this proportion, and "male-represented specialties" are defined as specialties with a female share that is lower than this proportion (see Appendix F for a list of specialties and their groupings). These female-represented specialties have been revealed to be perceived by Danish physicians as having a more balanced workload, being less competitive, and having more female role models (Korremann 1994). Plotting life cycle income trajectories for the two classes of occupations, panel C of Figure 3 illustrates how female-represented specialties are economically less favorable than male-represented specialties.

The quasi-experimental effects on this occupational choice are provided in panel B of Table 2. Indeed, we find that females are more likely to sort into their gender-represented specialty, where there is no longer-run effect on males. Since specialty choices in the residency determine the medical field of practice in the long run, this occupational sorting is consequential for physicians' career trajectories.

Overall, we have found, so far, that female physicians in the treatment group, as opposed to males, end up forgoing important human capital investments they would otherwise engage in, and they sort into economically less desirable stereotypical career paths at higher rates than they would otherwise choose. Together, these findings show that making less preferred early working-life choices results in important career outcomes that lead women to pursue disadvantaged career paths. Consequently, early-career experiences and choices can preserve and amplify underlying structures of gender bias in the labor market.

Unpacking the Treatment Bundle. Whereas our setting naturally involves a multi-dimensional treatment, as in most quasi-experiments, it has the advantage that we can investigate how the treatment may break down. Recall that we can conduct this analysis at two levels. In a first step, we can gauge the extent to which career effects can be attributed to a position's ranked quality versus the geographic labor market by comparing our treatment, middle, and control groups. In a second step, if the evidence points to a meaningful role for the geographic market, we can assess to which extent our characterization of local markets can account for the observed effects.

For the first step, we provide a decomposition of the effects we identified by calculating for each gender: (i) the "full" effect (treatment relative to control group); (ii) the "intermediate" effect (middle relative to the control group who primarily differ by position ranked quality); (iii) the difference between the "full" effect and the "intermediate" effect (capturing the difference between treatment and middle group who primarily differ by internship geographic market). We then calculate the share of the full effect that could be attributed to geographic market using the ratio of (iii) to (i).

These estimates are provided in panel A of Figure 4. Looking across the different outcomes, we find that there generally appears to be some role for the ranked quality of a position within a set of markets (as evidenced by the effects on the middle group). However, we find that the geographic labor market in which graduates begin their working lives seems to explain the bulk of the effect (as evidenced by the difference between the effects on the treatment and middle groups). As we have seen so far, males display no long-run effects on any of our outcomes across the various experimental groups. But for women, the decomposition attributes to the geographic market 74 percent of the effect on human capital investment, 100 percent of affiliation with university hospitals in the longer run, and 64 percent of sorting into female-represented specialties (as reported in the figure).

For the second step, having found evidence suggesting the geographic market matters, we assess the degree to which our two-dimensional characterization of the internship's locality (based on interning hospital type and degree of rurality) encompasses this composite. We use the "surrogate index" method (Athey et al. 2019). This method was proposed as a solution to the common challenge in estimating longer-term impacts of treatments, where outcomes of interest are observed with a long delay. The idea is to combine several shorter-term outcomes into the "surrogate index," which is the predicted value of the longer-term outcome given the shorter-term outcomes (the "surrogates") based on long-run observational data. Athey et al. (2019) show that the average treatment effect on the surrogate index equals the treatment effect on the long-term outcome. This is the case under the assumption that the long-term outcome is independent of the treatment conditional on the surrogate index, which forms the "surrogacy condition."[11]

We follow the implementation in Athey et al. (2019) and estimate the following statistical models. Let y_i represent the long-run outcome of interest, and let $s_i = (s_{1i}, \dots, s_{mi})$ be the vector of intermediate outcomes. To construct the surrogate index estimator, we use data on the control group to run the OLS regression:

$$y_i = \delta_0 + \sum_{j=1}^{m} \delta_j \times s_{ji} + \omega_{it}. \tag{3}$$

The surrogate index for the long-run outcome is the predicted value from this regression, which we denote by $\hat{y}_i$. We construct this index for each individual in our experimental sample by calculating: $\hat{y}_i = \hat{\delta}_0 + \sum_{j=1}^{m} \hat{\delta}_j \times s_{ji}$. The average treatment effect on the long-term outcome is estimated as the treatment effect of

[11] This method improves on previously suggested methods that use only one surrogate outcome in that it weakens the standard surrogacy condition for a single variable, with the notion that there is a greater likelihood that a set of intermediate outcomes could together satisfy the surrogacy condition.

the quasi-experiment on the predicted value of the long-term outcome. That is, we estimate the average treatment effect, β^s, using the following regression:

$$\hat{y}_i = \alpha^s + \beta^s \times Treat_i + \epsilon_{it}. \tag{4}$$

We bootstrap standard errors to account for estimation error from both stages of the surrogacy analysis.

We first apply the surrogacy index method to assess the extent to which our characterization of location can account for the effects within our ten-year analysis horizon of experimental data. To do so, we consider, as intermediate outcomes, the two dimensions of location we used: attachment to a university hospital in the internship period and whether the locality of the internship is rural. We use estimates from equation (4) for the effect on long-run outcomes that is based on the surrogacy index, β^s, and we compare it to the actual effect on long-run outcomes using our estimates from equation (2), β. The ratio β^s/β essentially evaluates the surrogacy condition in this application, which, in turn, gauges the degree to which the intermediate outcomes can explain the actual treatment effect.

Table 3 presents the findings. Across labor market outcomes, we find that the predicted treatment effects can account for the majority of the actual treatment effects we find for females. This supports the idea that our characterization of geographic markets (and the associated features we have shown) captures the bulk of the market composite. For males, the analysis predicts there should be non-negligible effects, whereas, in fact, we observe none. This suggests that males engage in actions in response to the lottery (such as those reflecting mobility) that mitigate the potential adverse effects of unfavorable internship allocations, which will guide us in the investigation of mechanisms underlying the gender divergence in long-run effects.

Predicted Long-run Earnings Gap. We conclude this subsection on labor market outcomes by discussing implications for earnings. In our context, while we have data on a long horizon of ten years, we could expect the effects on the career-defining labor market choices that we studied to translate into effects on earnings only later. This is because the returns to major human capital investments may materialize only in the very long run. In that sense, earnings could be insufficient for studying individuals' relative positions in the labor market in the analysis of early careers within the potentially lengthy period of human capital investments (what is commonly known as the "life-cycle bias," see Black and Devereux 2011). For example, we saw that the returns to a medical PhD materialize on average as late as 15 years after graduation (see Figure 3). Indeed, panel B of Table 2 displays no treatment effects on earnings in our analysis horizon of ten years.

Instead, we again take advantage of the surrogate index method. This is the exact scenario for which this tool had been developed; that is, for outcomes of interest (here earnings in the long run) that are observed with a very long delay. In the first step of estimating the "surrogate index," we include, as surrogates, all our labor market outcomes, and we use observational data on Danish physicians over three decades. Specifically, we include as intermediaries the year-ten position of the physician in the following outcomes: holding a PhD, having specialized in a female-represented specialty, bearing an affiliation with a university hospital, and residing in a less desirable location. We also include indicators for number of children since we find significant treatment effects on fertility in the next subsection. We then study the average treatment effect on the surrogate index for earnings predictions over the course of 30 years.

Panel B of Figure 4 presents the results. It displays, for different time horizons, the predicted treatment effects on earnings for males and females, as well as the predicted gender earnings gap. The predicted baseline gender gap grows over time, as the typical pattern in high-skill professions. In terms of treatment effects, there are no predicted long-run impacts on earnings for males as expected, reflecting the fact they displayed no effects on the entire array of labor market outcomes. In clear contrast, we find a widening effect for females. As there is no effect for males, this pattern in the effect on females that grows over time translates to a growing gender gap in earnings that is causally attributable to the lottery. Notably, our transitory treatment of variation in individuals' very first jobs alone can explain about 10-14 percent of the predicted long-run gender gap in physician earnings within the third decade after graduation.

6.2. Marriage Market Choices: Household Formation and Fertility

Family-related choices and career-related choices naturally intertwine. The postgraduation and early-career stages represent formative years with respect to family formation (Goldin and Katz 2008), and, in fact, the average age of graduates in our setting is 28.5. Appendix Figure E.1 first descriptively illustrates how partnership rates and fertility evolve among physicians' families after graduation. Indeed, the figures clearly show increased activity along these decision margins in the immediate years following graduation, specifically the internship period. Accordingly, we turn to investigate the interplay between the labor market and the marriage market by studying how early-career choices can causally affect family formation.

In the analysis that follows, we split our sample into two groups based on individuals' partnership status at the baseline pre-period: individuals who were partnered pre-lottery and individuals who were not partnered pre-lottery. Conceptually, the two subsamples differ with respect to the household decisions they face at the beginning of their careers. Partnered individuals enter the postgraduation planning period as a joint unit of two partners who make family planning choices. In comparison, single individuals additionally face a household formation choice through matching in the marriage market. Since their relevant pool of potential partners can be altered by the quasi-experiment, with its effects on graduates' environment and the set of people they may interact with (e.g., through changes to their workplaces and geography), we could expect implications for the matches formed by singles.

Partnership and Fertility. We begin by analyzing the sample of individuals who were single in the pre-period. For these individuals, partnership and fertility could both be important operative margins. Columns 1-2 of Table 4 summarize the effects on these outcomes. First, we find no detectable effects on the probability of having a partner. That is, for both genders, there is no difference between the treatment and control groups in the probability of having a partner in the longer run.

However, interesting patterns arise when we move on to studying fertility choices. We begin by looking at the longer-run impact of the lottery on an individual's number of children. Whereas there are no effects on men, women in treatment group exhibit an increase of 11.5 percent (=0.1422/1.2374) in the number of children in their families as a result of the internship placement variation. We then disentangle this result by studying the probability of having one child or more (i.e., becoming a parent) and the probability of having more than one child (i.e., higher fertility). The results show that the treatment effect is concentrated on having higher fertility: women in the treatment group are 7.1 pp more likely to have more than one child than women in the control group, which amounts to an effect of 16 percent.

In contrast, columns 3-4 of Table 4 show there are no such effects among individuals who were partnered in the pre-period. The lack of an impact on partnered individuals could suggest that the fertility effects on singles is less likely to be driven by an underlying shift in a household's family preferences due to the early labor market choices. Rather, the fertility effects could be related to differential matching in the marriage market among singles, which we investigate next.

Marriage Market Matching. We test the hypothesis of potential effects on matching patterns among single physicians by constructing measures of matching likelihoods based on a set of observables for the pool of their potential partners. The idea is that systematic differences for a given set of observables would be consistent with differential matching, as reflected by selection on observables.

The details of our analysis are as follows. We first take from the general population an approximate pool of individuals who could be potential partners for the subjects of our quasi-experiment. We select this pool based on gender and age. Specifically, for our single female physicians, we take the pool of all males of ages 30-50, and for our single male physicians, we take the pool of all females of ages 25-45.[12] We then predict, using lottery years 6-10, the match probability for each person i in the partner pool of marrying a person in our subject pool. We split the single subject pool on two dimensions: whether the physician belongs to the treatment group or the control group (indexed by $l \in \{t, c\}$), and whether the physician is male or female (indexed by $n \in \{m, f\}$). We then let $D_{i(l,n)}$ denote an indicator for a potential partner i marrying a physician from group (l, n), and we estimate a set of logistic regressions:

$$\Pr\left(D_{i(l,n)} = 1 | X_i\right) = F\left(\beta_{(l,n)} X_i\right),$$

[12] We choose these age ranges based on the empirical distribution of partners' ages in our experimental data in years 6-10 after graduation. We trim these age distributions at their 1st and 99th percentiles and round the resulting bounds to the nearest multiple of five for more intuitive ranges. For female physicians, the corresponding percentiles for the age of male partners are 29 and 51. For male physicians, the corresponding percentiles for the age of female partners are 26 and 45.

where $F(z) = \exp(z)/[1 + \exp(z)]$. In the observables vector, X_i, we include a third-order polynomial in age and whether the potential partner also holds a medical degree, and we additionally include year fixed effects. Finally, we calculate treatment/control ratios of the predictions:

$$\Pr\left(D_{i(t,n)} = 1|X_i\right)/\Pr\left(D_{i(c,n)} = 1|X_i\right), \quad (5)$$

for male/female subjects ($n \in \{m, f\}$). If we find that these ratios meaningfully deviate from 1, this evidence of selection on observables would imply differential matching across experimental groups; that is, it would provide evidence that the quasi-experiment has an effect on matching in the marriage market.

Panel A of Table 5 reports these ratios. For males, though highly precise due to the large potential partner pool, this ratio is economically very close to 1. It suggests no meaningful effects on males' marriage patterns, in line with the null effects on their number of children. However, consistent with the matching hypothesis we conjectured, we find a large deviation from 1 for our female subjects on the order of 25 percent. While the variation in placement does not affect whether you marry, it affects whom you marry. To investigate the direction in which the marriage patterns change due to the treatment, we analyze as outcomes the characteristics of the actual partners with whom our single subjects match. In panel B of Table 5, we find evidence of decreased assortative mating, whereby single women in the treatment group find older partners with larger age gaps and who are less likely to hold a medical degree.[13]

7. Mechanisms: What Can Explain the Gender Divergence?

We have found clear causal impacts on females' long-run labor market and marriage market choices in a direction that preserves and amplifies underlying structures of gender bias in the labor market. Our analysis cleanly identifies a specific source of variation: individuals' initial labor market sorting. It serves as a clean real-life laboratory that provides proof of concept for the far-reaching impacts of early-career choices and how they initiate significant divergence in long-run economic outcomes across gender. In this section, we support our main analysis with a characterization of the nature of the identified effects by investigating leading explanations for the gender divergence we have uncovered.

We focus on two candidate mechanisms that the literature has underscored as potential sources of gender disparities—family-career tradeoffs and workplace mentorship. We find strong evidence in support of these two mechanisms. We also investigate an alternative mechanism—that of differential preferences over entry-level positions. We find strong evidence against this explanation in our setting.

[13] The analysis in panel B of Figure 5 is descriptive and aims to characterize the causal effect we found on matching. It is endogenous since it includes only individuals who become partnered, though we have found no differential partnership rate across the experimental groups.

7.1. Family versus Career Tradeoff

A key aspect in the analysis of gender is that labor market and marriage market considerations could interact differentially for males and females (e.g., Goldin and Katz 2008, Bertrand et al. 2010, Kleven, Landais, and Søgaard 2019). Specifically, prior work has underscored that family responsibilities from the arrival of children hinder females' career advancement. Accordingly, we are interested in investigating whether such a mechanism could also serve as an explanation for the effects of early-career choices on the long-run divergence in outcomes.

A starting point that could hint in this direction is the underlying relative position of male and female physicians within their households. We consider the relative earnings of physicians out of the household earnings as a common proxy for labor market versus household production specialization. The plot in Appendix Figure E.2 shows that female physicians form more even households, suggesting that they may bear split responsibilities, and that males form households in which they are the primary earners, which may tilt their responsibilities away from the household and toward the labor market. This sets the scene for our key hypothesis that men and women differentially tackle the tension between family and career in their responses to initial sorting.

Human Capital and Fertility Interactions. Our identified differential effects on fertility by gender stand as initial evidence suggesting that the tension between family and career is a potential explanation. Indeed, our findings are consistent with the notion that women, unlike men, may crowd out long-run career goals by becoming more oriented toward the family as a result of unfavorable placements at the beginning of their working life. Such patterns are also consistent with the regularity that the more rural localities, which treatment group physicians sort into, display traditional gendered norms to a larger extent, as we have seen earlier (in panel C of Table 1). To directly test whether this tradeoff plays a role in practice, we ask: is it indeed the case that the women who have more children *also* invest less in human capital? Panel A of Table 6 finds clear evidence in support of this view, showing that single women in the treatment group exhibit an increase in the joint probability of not earning a PhD *and* having more than one child on the order of 7 pp (similar to the probability of having more than one child).

Proximity to Grandmothers. As the analysis pointed to the importance of location for successful labor market outcomes, we can further explore the family-career tradeoff hypothesis by studying location decisions that are either more family oriented or more career oriented. One natural relocation pattern that seems family oriented is the proximity to grandparents as potential childcare givers. In panel B of Table 6, we study, as an outcome, the distance to the physician's grandmother. The results show that, in the longer run, females in the treatment group locate closer to their mothers, with no effects on males. To assess whether these patterns could indeed link to family considerations, we study as an outcome the joint probability of living close to one's mother (based on a dummy for distance being below the sample mean) *and* having higher fertility (based on our measure of having more than one child that we found to be the active fertility margin). Closely in line with this hypothesis, we see that the effect of living closer to the grandmother is driven by women who were also induced by the treatment to have a higher fertility rate. In sum, only female physicians respond by locating closer to their mothers in the longer run, and this effect is concentrated among those who have more children due to the initial placement, suggesting females' location choices are tilted toward the family.

Search Behavior and Mobility. We have shown evidence suggesting that males may engage in career-oriented actions, in response to the lottery, that allow them to mitigate the potential adverse effects of unfavorable internship choices. A particular course of action we consider is search and mobility, which had been suggested in the classic work of Topel and Ward (1992) as driving males' early careers. Accordingly, we are interested in testing the conjecture that males and females may display different search behavior in the labor market in response to initial placements. One recently studied margin in the context of differential search behavior by gender is commuting distance (Le Barbanchon et al. 2021). In panel C of Table 6 (columns 1-2), we analyze the effect of the lottery on commuting distance. The evidence shows that, in contrast to females, males in the treatment group are more likely to commute further. This is consistent with the notion that differential willingness to commute could help mitigate adverse effects on males via differential responses to the treatment.

Another margin to investigate in relation to search is internal migration across labor markets. To this end, we consider the sample of pre-lottery singles who face the decision of forming a new household. We then study the physician's propensity to reside in the long run within the pre-lottery location of their new spouse. In panel C of Table 6 (columns 3-4), we find that households of male physicians show a decreased propensity to reside in the spouse's original location, revealing that men are more mobile and likely to migrate their households across labor markets. Together, the patterns suggest that male and female physicians exhibit differential priorities in their long-run location decisions, where females prioritize their family while males prioritize their careers.

7.2. Workplace and Mentorship

"Excess Sensitivity" to Workplace. To assess the role of employer-side factors, we begin with a general investigation of the potential role for workplace characteristics. The motivation for this analysis is the hypothesis that being differentially treated by the same entry-level employers (given there is a similar first stage across gender) could alter the career course of graduates differentially for males and females.[14] Specifically, we ask: do workplace "outcome-based" characteristics translate into employees' future

[14] We note that discrimination, in terms of unequal treatment of equally qualified interns of different gender, will fall within this part of the analysis. But we refrain from giving such interpretation to our findings in the absence of variation and data that could speak to discrimination explicitly.

outcomes differentially by gender? We analyze whether men and women display differential sensitivity in terms of their own subsequent placement (which is market based) when they are exposed, due to the lottery, to internship employers who do better or worse in placing their interns in subsequent positions.

The specific way in which we do this is by identifying hospital departments as employers and linking interns within an employer to the interns' next employer. An important dimension of the next position, for which we have consistent information that can be linked across hospital departments (as internship employers), interns, and their employers in the subsequent stage, is whether the next position is held at a university hospital. For each intern of a given gender, we calculate the leave-out-mean of how well their internship employer places its similar-gender interns in a university hospital later on. We refer to this measure as "employer intensity." Finally, we study interns' sensitivity to employer intensity by gender by regressing one's own probability of being employed at a university hospital in their next position on the intensity of their internship employer. The benchmark for the slope of full pass-through is one.

Panel A of Table 7 reports the results. We first show, as before, that males and females are similarly affected in the first stage, here in terms of how the quasi-experiment leads them to intern for employers who are "worse" in future placements. In contrast, we find that women's outcomes display a higher sensitivity than men's outcomes—by 30 percent more—to their employer's placement track record. Combining the patterns (of similar exposure but differential sensitivity) suggests that the gender divergence could be linked to being differentially treated by the market and how its characteristics ultimately translate to outcomes. It sheds light on how opportunities play out in practice and can shape into gender differentials in performance.

Mentorship. Having found that employer characteristics can matter, we turn to investigate the potential role of a particular workplace characteristic—the mentorship it provides. This is in light of the important literature that has suggested same-gender role models and mentors as a mechanism for gender inequalities in field of study and occupational choice.[15] Even more, this work has found strong influences of the gender of mentors on females, with little to no impact on males. Accordingly, we investigate whether variation from the treatment in terms of exposure to role models could provide an explanation in our setting.

We leverage information from the exit surveys of the later cohorts in our sample (after the system was digitized) for whom we have exact linkages, with full names, between interns, their formally assigned mentor, and the head of the education program at the hospital department they intern. The names allow us to impute (with error) the gender of these role models (see details of this procedure in Appendix G).

Panel B of Table 7 (columns 1-4) shows that the quasi-experiment leads to a large decline in exposure to female role models (for both males and females), as captured by a decreased probability of

[15] See, for example, Bettinger and Long (2005), Carrell et al. (2010), Blau et al. (2010), Dennehy and Dasgupta (2017), Kofoed and McGovney (2019), Porter and Serra (2020), and Ginther et al. (2020).

having been assigned a female formal mentor or program chair in the internship. Moreover, we further take advantage of the internship exit surveys to study how interns translate this variation into their perceived experience of the mentorship. The exit surveys have a dedicated section for interns to evaluate their mentor in terms of the training plan, provision of feedback, and advising on professional and career development (see Appendix G). We find that females in the treatment group rate this aspect of the internship meaningfully lower (by 0.25 of a standard deviation), whereas there is no detectable effect on males (panel B of Table 7, columns 5-6). The evidence is consistent with the notion of variation in mentorship as an operative mechanism.

7.3. Alternative Explanation: Preferences over Entry-Level Positions

There is an important discussion in the gender literature about whether gender differences in economic choices, such as college majors and occupations, might actually stem from differences in preferences (Bertrand 2020). We can test this hypothesis in our context by leveraging information on preferences over the entry-level local labor markets and internship specialties.

We proceed in two steps. First, we utilize our measure for market desirability, which reveals students' location preferences through their lottery-based choices. We now construct these market rankings (based on the average lottery rank of the interns who choose to sort into it) separately for males and females and compare across them. Panel A of Figure 5 illustrates the results. Each dot represents a local labor market, where the x-axis denotes male rankings and the y-axis denotes female rankings. We plot the fitted line and report its slope, where the benchmark of non-differential ranking by gender is one (the 45-degree line). Overall, the estimation is notably close to the benchmark case under the null that males and females have similar average priorities over the entry-level markets.[16] Hence, we find no evidence of differential preferences over entry-level market locations in our setting. Second, we investigate graduates' occupational preferences in their entry-level jobs. We base our analysis on revealed preferences for specialty choices in internships. Within the primary positions at hospitals, interns can broadly choose between internal medicine and surgery, and, within the secondary positions in primary care, interns can choose between general medicine and psychiatry. For each gender, we split the sample by deciles according to lottery ranks. Then, for each of the two types of positions, we calculate, over deciles, the gender-specific cumulative probability of making a particular choice of specialty over the other. We plot in panels B and C of Figure 5 the gender-specific CDFs against one another, where the 45-degree line serves as a benchmark when preferences of specialties are similar across gender. Again, we do not find any systematic differences across gender in these choices either. Put together, the evidence strongly suggests that (ex-ante) preferences over entry-level

[16] We reach a similar conclusion if we instead use the information we have for the earlier cohorts about students' binding pre-placement rankings of all local labor markets, as reported in their priority lists (see panel D of Appendix Figure C.1).

positions are not driving our results. Of course, it is still possible that preferences change ex-post differentially across gender in response to the treatment.

8. Conclusion

Using a randomized lottery that determines Danish physicians' entry-level placements, we identify significant impacts of early labor market sorting on longer-run career versus family choices. These far-reaching effects encompass human capital investment, occupational choice, family formation, and fertility. We find that the long-run effects are entirely concentrated among females, thereby providing evidence of a novel route that initiates and perpetuates gender inequality and gender-biased labor market norms. We find strong evidence for gender differences in making choices that differentially prioritize careers over the family and the workplace's mentorship as channels that can explain the long-run gender divergence.

Our analysis highlights how persistent gender inequality can arise even in an institutionally equitable setting. As such, our findings imply that policies that aim to achieve outcome-based gender equality cannot only rely on leveling the starting playing field. Rather, such policies should target the ways in which these opportunities play out in practice and shape into gender-differential choices over the course of the formative years of early-career stages.

Relatedly, our findings can also have implications for the efficacy of actual efforts that have taken place in the form of lotteries with the direct aim of reducing disparities and biases. In the research funding domain, for example, foundations are increasingly utilizing lotteries in the allocation of funds in the direct effort to reduce funding biases, including public funds (such as the Swiss National Science Foundation) and large private foundations (such as Novo Nordisk Foundation and Volkswagen Foundation). Extending the evidence we show here, such efforts could induce unintended consequences in other dimensions, such as disparities in how long-run returns play out, and prove surprisingly less effective than hoped for overall.

References:

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

Alsan, Marcella, Owen Garrick, and Grant Graziani (2019). Does Diversity Matter for Health? Experimental Evidence from Oakland. *American Economic Review*, 109 (12): 4071-4111.

Arora, A., Goff, L., and Hjort, J. (2021). Pure-Chance Jobs versus a Labor Market: The Impact on Careers of a Random Serial Dictatorship for First Job Seekers. In *AEA Papers and Proceedings*, 111: pp. 470-75.

Athey, S., Chetty, R., Imbens, G.W., and Kang, H. (2019). The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely. NBER Working Paper 26463.

Azmat, G., Calsamiglia, C., and Iriberri, N. (2016). Gender Differences in Response to Big Stakes. *Journal of the European Economic Association*, 14(6), 1372–1400.

Azmat, G., and Ferrer, R. (2017). Gender Gaps in Performance: Evidence from Young Lawyers. *Journal of Political Economy*, 125(5), 1306–1355.

Bhattacharya, J., Hyde, T., and Tu, P. (2013). *Health Economics*. Macmillan International Higher Education.

Bettinger, E.P., and Long, B.T. (2005). Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students. *American Economic Review*, *95*(2), 152–157.

Bertrand, M. (2011). New Perspectives on Gender. In *Handbook of Labor Economics*, 4, 1543–1590.

Bertrand, M. (2020). Gender in the Twenty-First Century. *AEA Papers and Proceedings*, 110, 1-24.

Bertrand, M., Goldin, C., and Katz, L. F. (2010). Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. *American Economic Journal: Applied Economics*, 2(3), 228–255.

Black, S.E., and Devereux, P.J. (2011). Recent Developments in Intergenerational Mobility. In *Handbook of Labor Economics*, 4, 1487–1541.

Blau, F. D., and Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55(3), 789–865.

Blau, F. D., Currie, J. M., Croson, R. T., and Ginther, D. K. (2010). Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial. *American Economic Review*, 100(2), 348–52.

Brotherton, S. E., and Etzel, S. I. (2018). Graduate Medical Education, 2017-2018. *JAMA*, 320(10), 1051–1070.

Bursztyn, L., Fujiwara, T., and Pallais, A. (2017). 'Acting Wife': Marriage Market Incentives and Labor Market Investments. *American Economic Review*, 107(11), 3288–3319.

Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, Competitiveness, and Career Choices. *Quarterly Journal of Economics*, 129(3), 1409–1447.

Buser, T., and Yuan, H. (2019). Do Women Give Up Competing More Easily? Evidence from the Lab and the Dutch Math Olympiad. *American Economic Journal: Applied Economics*, 11(3), 225–252.

Cai, X., Lu, Y., Pan, J., and Zhong, S. (2019). Gender Gap under Pressure: Evidence from China's National College Entrance Examination. *Review of Economics and Statistics*, 101(2), 249–263.

Caliendo, M., Lee, W. S., and Mahlstedt, R. (2017). The Gender Wage Gap and the Role of Reservation Wages: New Evidence for Unemployed Workers. *Journal of Economic Behavior and Organization*, 136, 161–173.

Card, D., Cardoso, A. R., and Kline, P. (2016). Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women. *Quarterly Journal of Economics*, 131(2), 633–686.

Carrell, S.E., Page, M.E., and West, J.E. (2010). Sex and Science: How Professor Gender Perpetuates the Gender Gap. *Quarterly Journal of Economics*, 125(3), 1101–1144.

Cheng, Stephanie D. (2020). Careers Versus Children: How Childcare Affects the Academic Tenure-Track Gender Gap.

Chetty, R., and Hendren, N. (2018). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. *Quarterly Journal of Economics*, 133(3), 1107–1162.

Cortés, P., Pan, J., Reuben, E., Pilossoph, L. and Zafar, B. (2022). Gender Differences in Job Search and the Earnings Gap: Evidence from the Field and Lab.

Cullen, Z. and Perez-Truglia, R. (2019). The Old Boys' Club Schmoozing and the Gender Gap. NBER Working Paper No. w26530.

Damm, A. P., and Dustmann, C. (2014). Does Growing Up in a High Crime Neighborhood Affect Youth Criminal Behavior? *American Economic Review*, 104(6), 1806–32.

Danish Economic Councils (2015). Dansk Økonomi, Forår 2015.

Danish Ministry of Health (1989). Lægers Kliniske Videreuddannelse og arbejdstilrettelæggelse, betænkning nr. 1183, September 1989.

DellaVigna, S., Heining, J., Schmieder, J.F. and Trenkle, S. (2022). Evidence on Job Search Models from a Survey of Unemployed Workers in Germany. *Quarterly Journal of Economics*, 137(2): pp. 1181-1232.

Dennehy, T.C., and Dasgupta, N. (2017). Female Peer Mentors Early in College Increase Women's Positive Academic Experiences and Retention in Engineering. *PNAS*, 114(23), 5964–5969.

Eckert, F., Walsh, C., and Hejlesen, M. (2019). The Return to Big City Experience: Evidence from Danish Refugees.

Exley, C. L., and Kessler, J. B. (2019). The Gender Gap in Self-Promotion.

Field, E., Jayachandran, S., Pande, R., and Rigol, N. (2016). Friendship at Work: Can Peer Effects Catalyze Female Entrepreneurship? *American Economic Journal: Economic Policy*, 8(2), 125–53.

Finkelstein, A., Gentzkow, M., and Williams, H. (2016). Sources of Geographic Variation in Health Care: Evidence from Patient Migration. *Quarterly Journal of Economics*, 131(4), 1681–1726.

Ginther, D. K., Currie, J. M., Blau, F. D., and Croson, R. T. (2020). Can Mentoring Help Female Assistant Professors in Economics? An Evaluation by Randomized Trial. *AEA Papers and Proceedings*, 110, 205–09.

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *Quarterly Journal of Economics*, 118(3), 1049–1074.

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, 104(4), 1091–1119.

Goldin, C. (2021). Career and Family: Women's Century-Long Journey toward Equity. Princeton University Press.

Goldin, C., and Katz, L. F. (2008). Transitions: Career and Family Life Cycles of the Educational Elite. *American Economic Review*, 98(2), 363–69.

Holman, L., Stuart-Fox, D., and Hauser, C. E. (2018). The Gender Gap in Science: How Long until Women are Equally Represented? *PLoS biology*, 16(4), e2004956.

Huang, J., Gates, A. J., Sinatra, R., and Barabási, A. L. (2020). Historical Comparison of Gender Inequality in Scientific Careers across Countries and Disciplines. *PNAS*, 117(9), 4609–4616.

Iriberri, N., and Rey-Biel, P. (2019). Competitive Pressure Widens the Gender Gap in Performance: Evidence from a Two-stage Competition in Mathematics. *Economic Journal*, 129(620), 1863–1893.

Kirkeboen, L. J., Leuven, E., and Mogstad, M. (2016). Field of Study, Earnings, and Self-Selection. *Quarterly Journal of Economics*, 131(3), 1057-1111.

Kleven, H., Landais, C., Posch, J., Steinhauer, A., and Zweimuller, J. (2019). Child Penalties across Countries: Evidence and Explanations. In *AEA Papers and Proceedings*, 109, 122–26.

Kleven, H., Landais, C., and Søgaard, J. E. (2019). Children and Gender Inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4), 181–209.

Kofoed, M. S., and McGovney, E. (2019). The Effect of Same-Gender or Same-Race Role Models on Occupation Choice: Evidence from Randomly Assigned Mentors at West Point. *Journal of Human Resources*, 54(2), 430–467.

Korremann, G. (1994), Læger og køn – spiller det en rolle? Nordisk Ministeråd, København, *TemaNord 1994:597*

Laird, J. and Nielsen, T. (2016). The Effects of Physician Prescribing Behaviors on Prescription Drug Use and Labor Supply: Evidence from Movers in Denmark.

Le Barbanchon, T., Rathelot, R., and Roulet, A. (2021). Gender Differences in Job Search: Trading off Commute Against Wage. *Quarterly Journal of Economics*, 136(1), 381-426.

Lundberg, S., and Stearns, J. (2019). Women in Economics: Stalled Progress. *Journal of Economic Perspectives*, 33(1), 3–22.

National Board of Health (2009). Sundhedsstyrelsens målbeskrivelse for den kliniske basisuddannelse.

Niederle, M. and Vesterlund, L. (2007). Do Women Shy Away from Competition? Do Men Compete Too Much? *Quarterly Journal of Economics*, 122(3), 1067–1101.

OECD (2012). Health at a Glance 2011: OECD Indicators 2011. OECD Publishing, Paris.

Olivetti, C. and Petrongolo, B. (2016). The Evolution of Gender Gaps in Industrialized Countries. *Annual Review of Economics*, 8(1), 405–434.

Ono, T., Schoenstein, M., and Buchan, J. (2014). Geographic Imbalances in Doctor Supply and Policy Responses. OECD Health Working Paper.

Porter, C., and Serra, D. (2020). Gender Differences in the Choice of Major: The Importance of Female Role Models. *American Economic Journal: Applied Economics*, 12, 226–254.

Rubinstein, Y., and Weiss, Y. (2006). Post Schooling Wage Growth: Investment, Search and Learning. *Handbook of the Economics of Education*, 1, 1–67.

Sarsons, H. (2019). Interpreting Signals in the Labor Market: Evidence from Medical Referrals.

Simoens, S., and Hurst, J. (2006). The Supply of Physician Services in OECD Countries. OECD Health Working Paper.

Sloan, F.A., and Edmunds, M. eds. (2012). Geographic Adjustment in Medicare Payment: Phase I: Improving Accuracy. *National Academies Press.*

Topel, R. H., and Ward, M. P. (1992). Job Mobility and the Careers of Young Men. *Quarterly Journal of Economics*, 107(2), 439–479.

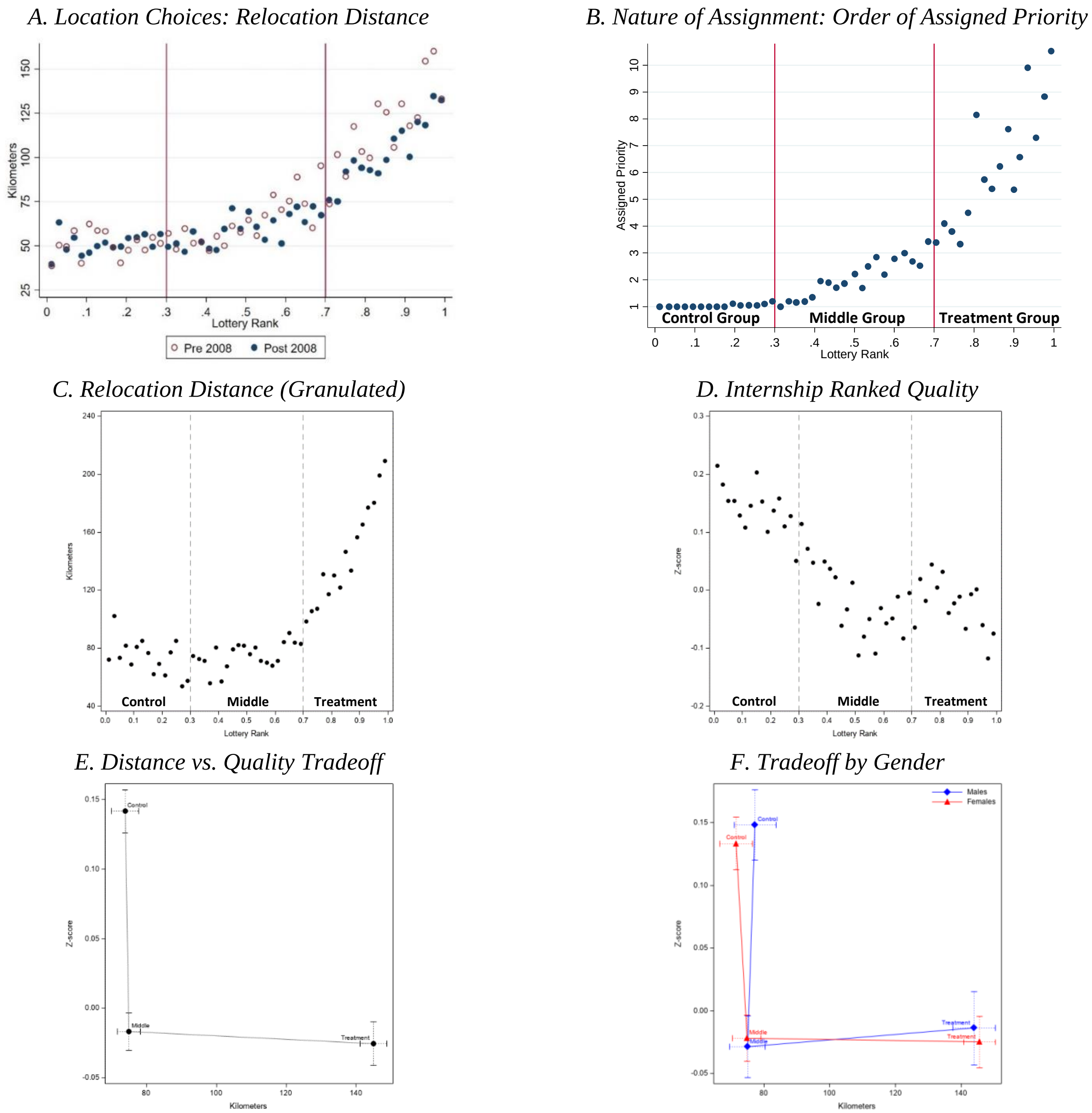
von Wachter, Till. (2020). The Persistent Effects of Initial Labor Market Conditions for Young Adults and Their Sources. *Journal of Economic Perspectives*, 34(4), 168–194.

Wasserman, M. (forthcoming). Hours Constraints, Occupational Choice, and Gender: Evidence from Medical Residents. *Review of Economic Studies*.

Weiss, Y. (1986). The Determination of Life Cycle Earnings: A Survey. *Handbook of Labor Economics*, 1, 603–640.

Zeltzer, D. (2020). Gender Homophily in Referral Networks: Consequences for the Medicare Physician Earnings Gap. *American Economic Journal: Applied Economics*, 12, 169–197.

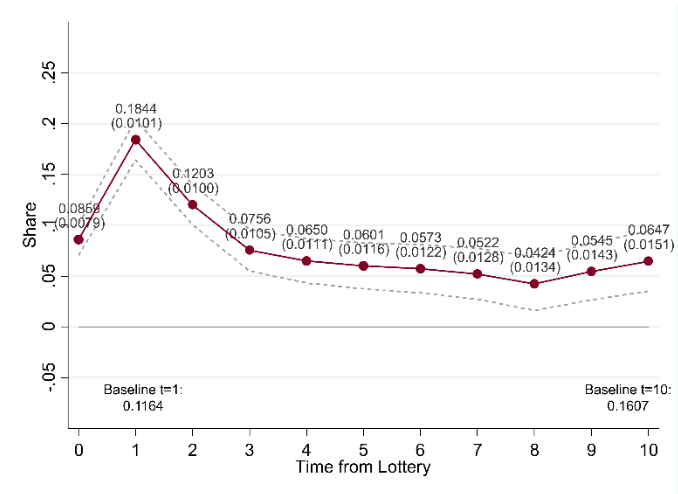
Figure 1: Patterns of Internship Choices and the First Stage



Notes: These figures study the nature of placement to internships. In panel A, we calculate, for each student, the distance (in kilometers) between their municipality of residence at the time of the lottery and their municipality of the internship, which captures their "relocation distances." This figure plots a graduating student's relocation distance against the student's lottery rank, where we split cohorts around 2008, when the process was digitized. Internship location is based on the physician's workplace in period one (at the municipality level), as reported in annual employment registers as of the month of November. In panel B, we use the rankings of all local labor markets that had been solicited among the earlier cohorts as part of the allocation process. We plot individuals' pre-placement rankings of the local labor market they were assigned to (where 1 is highest priority) against the percentile ranks of their lottery numbers (within their graduating cohort). Panels C-E investigate the relationship between relocation distance and an internship's ranked quality against lottery rank using information from after the system was digitized, where data on both dimensions are linked to interns. Panel C measures the relocation distance using more precise information on the internship location (that of the particular location) that is reported directly for later cohorts. Panel D measures ranked quality at the hospital-department level, using the leave-one-out mean of the overall evaluation normalized by the overall mean and standard deviation of this measure (to create a z-score). Panel E aggregates the information from panels C and D across our "control" group, "treatment" group, and "middle" group, and it plots the averages of the two dimensions simultaneously for each group. Panel F replicates panel E, split by gender. Panels E-F also display the corresponding 95-percent confidence intervals.

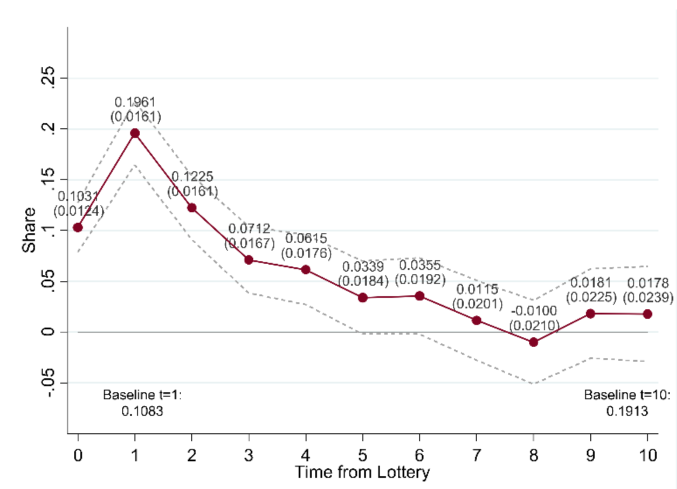
Figure 2: Dynamics of Geographic Sorting

A. Effects on Overall Sample

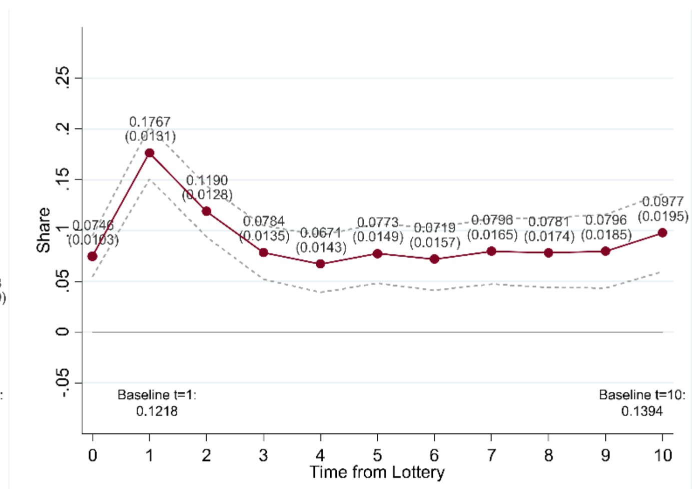


B. Effects by Gender

Males



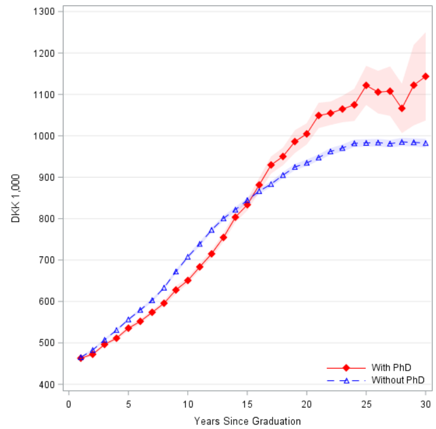
Females



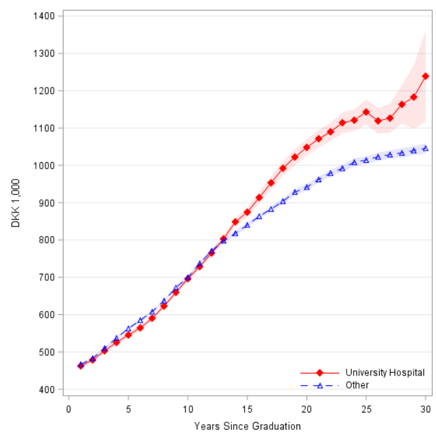
Notes: This figure plots the dynamic effects of the lottery on the probability of sorting into less desirable local labor markets. We plot the β_τ estimates from equation (1) for years 0 to 10, along with their 95-percent confidence intervals, where the x-axis denotes the year relative to the lottery. The early years (0-1) mechanically capture the first stage effect on the internship position, and the later years (6-10) capture the longer-run impact on households' geographic location decisions. Panel A includes the overall sample, and panel B splits the sample by gender.

Figure 3: Life Cycle Income Trajectories

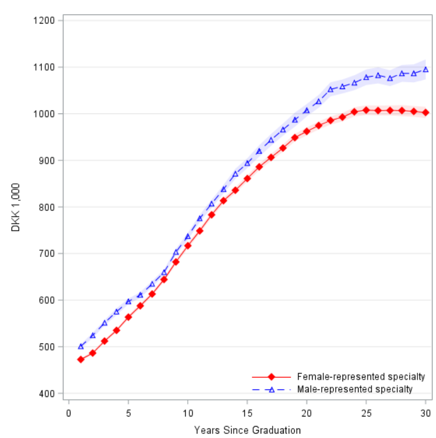
A. Medical PhD



B. Affiliation with University Hospitals



C. Gender-Represented Specialties

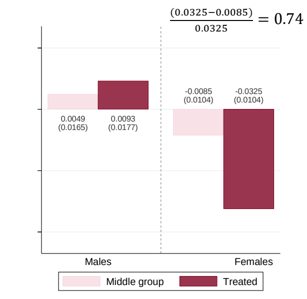


Notes: These figures plot income paths by years since graduation for the sample of all Danish physicians. Shaded areas represent 95-percent confidence intervals. We use a comprehensive measure of income from any source, including pre-tax wage earnings, capital income, government transfers, and self-employment business revenues. The figures compare the income trajectories of physicians who hold or do not hold a medical PhD (panel A), are affiliated or are not affiliated with a university hospital (panel B), and practice medicine in a female-represented specialty or in a male-represented specialty (panel C).

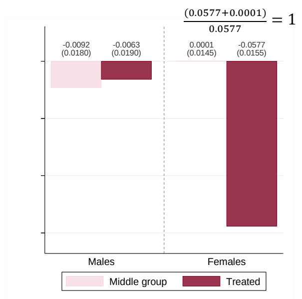
Figure 4: Labor Market Outcomes

A. Unpacking the Treatment Bundle

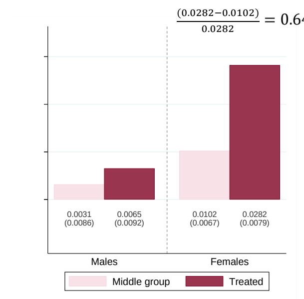
Medical PhD



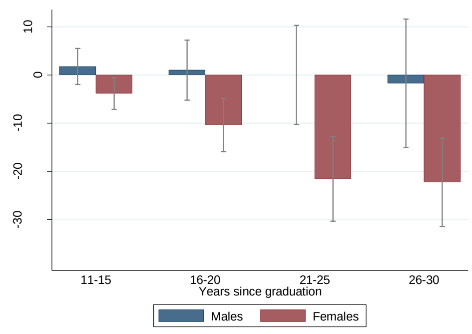
Affiliation with University Hospitals



Gender-Represented Specialties



B. Predicted Long Run Treatment Effects on Earnings



Years since graduation		***11-15***	***16-20***	***21-25***	***26-30***
Predicted Effect:	Males	1,742	1,004	-13	-1,739
	Females	-3,837	-10,415	-21,597	-22,265
Predicted baseline gender gap		-93,259	-119,519	-151,317	-219,266

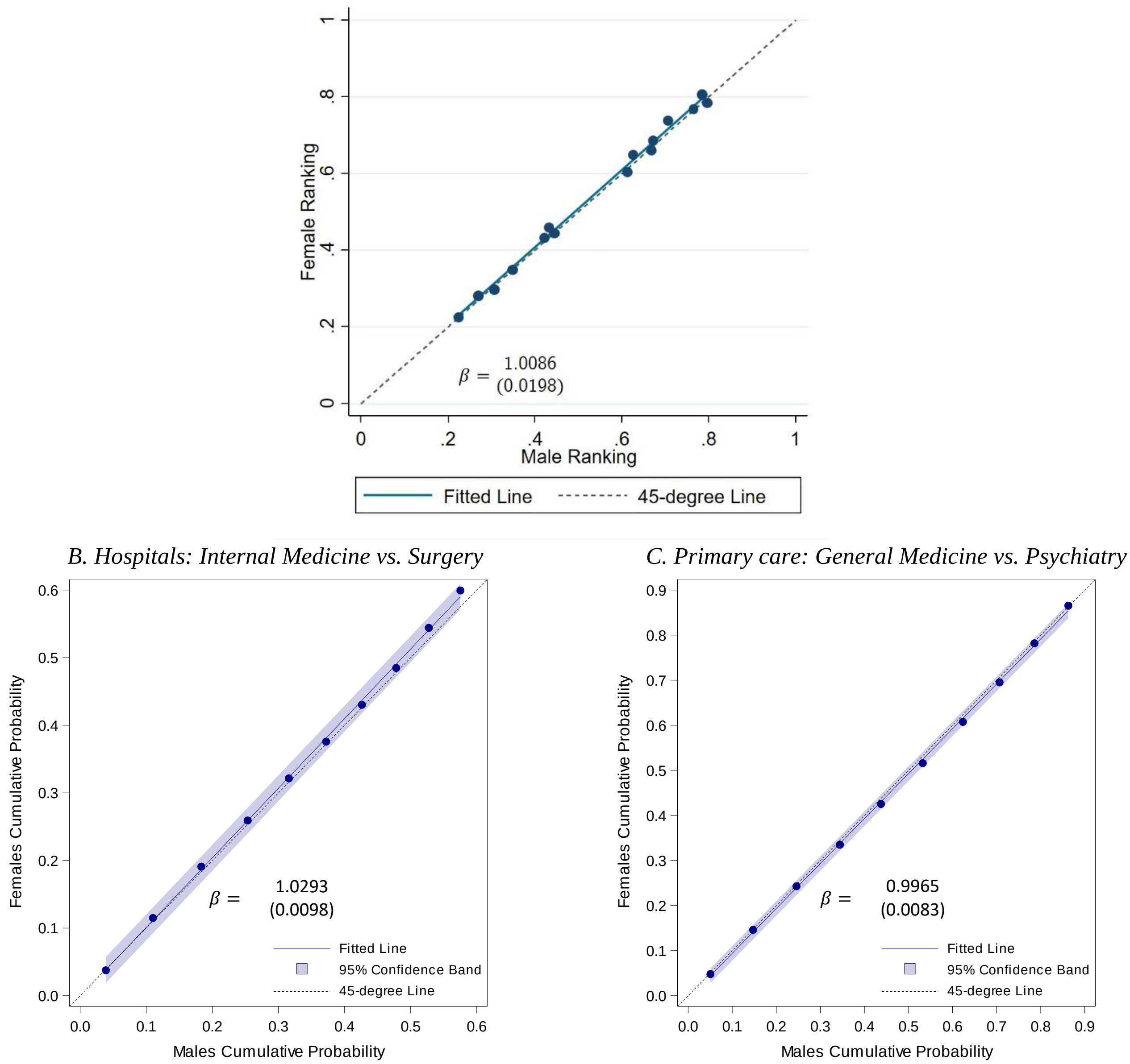
Notes: This figure investigates the longer-run effects on labor market outcomes. Panel A unpacks the overall treatment effect by providing the treatment effect on the "treatment" group (as compared to the "control" group) and the treatment effect on the "middle" group (as compared to the "control" group). These are estimated using a modification to equation (2), where we also include the middle group of the internship lottery and add an indicator for belonging to that group. The corresponding regression estimates are presented in Appendix Table E.1 (column 3). For each of the studied outcomes, we also provide a calculation of the share of the difference in treatment effects across the treatment group and the middle group out of the full effect on the treatment group. Panel B plots the estimates of the surrogate index predictions of the effects on long-run earnings. It is constructed in two steps. In the first step, we use the entire sample of physicians from 1980-2016 and regress (separately for males and females) earnings in years 11-15, 16-20, 21-25, and 26-30 after graduation from medical school on characteristics of the physicians ten years after graduation. These characteristics include indicators of the individual's values ten years after graduation for the following outcomes: having a medical PhD, employment at a university hospital, having a gender represented specialty, residing in a less desirable location, having no children, having one child, and having three children or more (where the omitted category is having two children). We use these regressions to construct the surrogate index for each person, that is, the predicted value for earnings in the long run. In the second step, we use the predicted values from the first step to estimate the effects on the surrogate index using the experimental sample. The estimates and their 95-percent confidence intervals are plotted in the graph. Standard errors are bootstrapped to account for estimation error from the two steps of the surrogacy analysis. The corresponding point estimates for the predicted effects are reported in the table below the figure, including the predicted baseline gender gap in earnings.

Figure 5: Preferences over Local Labor Markets and Medical Specialties by Gender

A. Location Preferences

B. Hospitals: Internal Medicine vs. Surgery

C. Primary care: General Medicine vs. Psychiatry



Notes: This figure compares male and female graduates' revealed preferences over entry-level local labor markets and internship specialties. Panel A investigates preferences over local labor markets. We use our measure for market desirability that reveals students' preferences through their lottery-based choices. We construct these market rankings based on the average lottery rank of the interns who choose to sort into it, separately for males and females, and compare across them. Each dot represents a local labor market, where the x-axis denotes male rankings and the y-axis denotes female rankings. We plot the fitted line, as well as the 45-degree line, which is the benchmark under non-differential rankings by gender, and we also report the slope of the fitted line, where the benchmark of non-differential ranking is one. Panels B-C investigate preferences over internship specialties. Within the primary positions at hospitals, interns can choose between internal medicine and surgery, and, within the secondary positions in primary care, interns can choose between general medicine and psychiatry. For each gender, we split the sample by deciles according to lottery ranks. Then, for each of the two types of positions, we calculate, over deciles, the gender-specific cumulative probability of making a particular choice of specialty over the other. We plot the gender-specific CDFs against one another, where, again, the 45-degree line serves as a benchmark when preferences of specialties are similar across gender. We also plot the fitted line along with 95-percent confidence intervals and report its slope.

Table 1: Characterization of Locations

A. Characteristics of Less Desirable Labor Markets

	University Hospital (1)	Rural Location (2)
Less Desirable Labor Market	-0.3064***	0.6153***
	(0.0796)	(0.1451)
Constant	0.4034***	0.0130
	(0.0581)	(0.1059)
Counties	15	15

B. Characteristics of University Hospitals

		Non-University	University	Difference	*p*-value
Scale	Unique Patients	42,403	86,437	44,034***	<0.0001
	Admissions	82,741	160,380	77,639***	<0.0001
	Procedures	28,947	64,450	35,503***	0.0001
Technology	Unique Procedures	816	1,485	669***	<0.0001
	CT Scanner (probability)	0.75	0.98	0.23***	0.0002
	CT Scans	12,839	41,472	28,633**	0.0202
	MRI Scanner (probability)	0.58	0.98	0.41***	<0.0001
	MRI Scans	6,380	24,678	18,298***	0.0018
Human Capital	Medical Specialties	9.9	16.5	6.6***	<0.0001
	Specialists with PhD (share)	0.069	0.156	0.087***	<0.0001

C. Characteristics of Rural Locations

		Urban	Rural	Difference	*p*-value
Demographics	Population density (capita per sq km)	1,681	83	-1,598***	0.0045
	Population size (capita)	165,284	53,849	-111,435***	0.0027
	College degree (%, ages 25-64)	32.8	20.3	-12.5***	<0.0001
	DI recipients (%, ages 17-64/66)	5.9	8.4	2.5***	<0.0001
	Annual income (DKK, ages 25-59)	396,200	349,271	-46,929***	<0.0001
Health and Healthcare	Primary care expenditure per capita (DKK)	450	617	167***	<0.0001
	Hospital visits per capita	0.84	0.97	0.13***	<0.0001
	Daily smokers, %	16.3	18.7	2.4***	<0.0001
Amenities and Norms	Home prices per square meter (DKK)	15,674	7,484	-8,190***	<0.0001
	Revenue from income tax per capita (DKK)	39,352	36,087	-3,265***	0.0041
	Places in daycare (%, ages 0-2)	40.0	22.7	-17.4***	<0.0001
	Expenditure on culture, sports, and leisure (per capita)	1,693	1,477	-216***	0.0019
	Women elected officials (%)	34.2	27.8	-6.3***	0.0003
	Parental leave, males (z-score)	0.023	-0.066	-0.089***	0.0003
	Parental leave, females (z-score)	-0.017	0.054	0.071***	<0.0001

Notes: Panel A characterizes the degree to which the desirability of local labor markets (counties) is predictive of the probability of interning in a university hospital or in a rural municipality. We run regressions at the county level, where the available data collapses two counties within the capital of Copenhagen into one. Standard errors are reported in parentheses. Panel B provides characteristics of hospitals (with a total of 51 nationally), split by whether they are non-university or university hospitals. We use data from the national patient register, the registries for income and education, and the authorization register. Panel C provides characteristics of rural versus urban municipalities, where the classification follows the formal definitions used by the Danish Economic Councils (2015). We use data from: "Municipal Key Figures," Ministry of Interior Affairs and Housing ("Kommunale Nøgletal," Indenrigs- og boligministeriet); "Housing Market Statistics," Finance Denmark ("Boligmarkedsstatistikken," Finans Danmark); "National Goals," Ministry of Health ("Nationale mål," Sundhedsministeriet); and the absence, income, population, education, national health insurance, and national patient registers. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 2: Labor Market Outcomes

A. Dynamics in Human Capital Investment: Obtaining a Medical PhD

	Males (1)	Females (2)
Treatment Effect at *t*		
6	0.0053	-0.0125*
	(0.0146)	(0.0067)
7	0.0171	-0.0295***
	(0.0186)	(0.0105)
8	0.0040	-0.0330**
	(0.0223)	(0.0146)
9	0.0094	-0.0444**
	(0.0251)	(0.0178)
10	0.0055	-0.0542***
	(0.0282)	(0.0209)
Counterfactual at t = 10	0.2697	0.2131
Average Treatment Effect	0.0093	-0.0325***
	(0.0177)	(0.0104)
Constant	0.1711***	0.1121***
	(0.0124)	(0.0080)
Individuals	1,551	2,306

B. Longer Run Effects

	Affiliation with University Hospitals		*Gender-Represented Specialties*		*Earnings*	
	Males	Females	Males	Females	Males	Females
Average Treatment Effect	-0.0063	-0.0577***	0.0065	0.0282***	11,750	1,135
	(0.0190)	(0.0155)	(0.0092)	(0.0079)	(9,110)	(5,802)
Constant	0.4627***	0.4448***	0.0706***	0.0743***	664,811***	543,388***
	(0.0134)	(0.0111)	(0.0063)	(0.0050)	(6,355)	(3,969)
Individuals	1,830	2,771	1,706	2,544	1,674	2,521
Treatment at t=10	0.0026	-0.0647***	0.0090	0.0578***	15,538	8,666
	(0.0277)	(0.0226)	(0.0238)	(0.0224)	(15,668)	(10,742)
Constant	0.4052***	0.4050***	0.1931***	0.2459***	747,040***	600,326***
	(0.0198)	(0.0162)	(0.0169)	(0.0153)	(11,378)	(7,054)
Individuals	1,262	1,824	1,123	1,586	1,024	1,528

Notes: This table reports effects of early-career choices on labor market outcomes. Panel A studies the dynamics in human capital investment using, as an outcome, an indicator for the completion of medical PhD. It provides estimates for β_τ using equation (1), starting from year 6, which is when PhD completion begins to materialize following graduation from medical school. Counterfactuals are calculated as averages of the control group's outcomes. Panel B studies the long-run effects of the lottery on different labor market outcomes. Estimates are based on equation (2), where, for the average treatment effects, we include years 6-10 after graduation. We study as outcomes the probability of being affiliated with a university hospital, the probability of sorting into gender-represented specialties, and earnings, which are winsorized at their 99th percentile. Robust standard errors, clustered at the individual level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 3: Labor Market Outcomes—Unpacking the Treatment Bundle

	Males			Females		
	Baseline Projection (1)	Predicted Effect (2)	Actual Effect (3)	Baseline Projection (4)	Predicted Effect (5)	Actual Effect (6)
A. Obtaining a Medical PhD						
University	0.0711***			0.0386**		
Hospital (t=1)	(0.0247)			(0.0164)		
Rural (t=1)	-0.0496			-0.0549**		
	(0.0482)			(0.0220)		
Treatment		-0.0317***	0.0093		-0.0215***	-0.0325***
		(0.0053)	(0.0177)		(0.0038)	(0.0104)
Constant	0.1348***	0.1711***	0.1711***	0.0931***	0.1121***	0.1121***
	(0.0171)	(0.0068)	(0.0124)	(0.0124)	(0.0047)	(0.0080)
Observations	3,104	6,337	6,337	4,569	9,157	9,157
Individuals	765	1,551	1,551	1,145	2,306	2,306
F-statistic	5.36			8.80		
p-value	0.0049			0.0002		
B. Affiliation with University Hospitals						
University	0.0559**			0.1026***		
Hospital (t=1)	(0.0278)			(0.0228)		
Rural (t=1)	-0.1115			-0.1819***		
	(0.0684)			(0.0424)		
Treatment		-0.0321***	-0.0063		-0.0591***	-0.0577***
		(0.0067)	(0.0190)		(0.0031)	(0.0155)
Constant	0.4359***	0.4627***	0.4627***	0.3957***	0.4448***	0.4448***
	(0.0219)	(0.0070)	(0.0134)	(0.0182)	(0.0021)	(0.0111)
Observations	3,834	7,752	7,752	5,726	11,510	11,510
Individuals	915	1,830	1,830	1,373	2,771	2,771
F-statistic	4.23			26.01		
p-value	0.0148			0.0000		
C. Gender-Represented Specialties						
University	-0.0066			-0.0332***		
Hospital (t=1)	(0.0133)			(0.0103)		
Rural (t=1)	-0.0053			0.0856***		
	(0.0329)			(0.0253)		
Treatment		0.0019	0.0065		0.0221***	0.0282***
		(0.0036)	(0.0092)		(0.0014)	(0.0079)
Constant	0.0745***	0.0706***	0.0706***	0.0884***	0.0743***	0.0743***
	(0.0104)	(0.0043)	(0.0063)	(0.0086)	(0.0009)	(0.0050)
Observations	3,468	7,045	7,045	5,142	10,325	10,325
Individuals	846	1,706	1,706	1,263	2,544	2,544
F-statistic	0.12			13.52		
p-value	0.8829			0.0000		

Notes: This table provides surrogate index analyses for our labor market outcomes, separately for males and females, to unpack the multi-dimensional quasi-experiment. Within each gender, the first column regresses the outcome of interest in years 6-10 on indicators of interning at a university hospital and in a rural location. We estimate this relationship using the sample of unconstrained interns from our control group. We use these regressions to construct the predicted values of the outcomes in years 6-10, that is, the "surrogate index." The second column, within each gender, then regresses the surrogate index on the treatment status using our subject pool to provide the predicted treatment effect of the internship lottery. In this estimation, standard errors are bootstrapped to account for the estimation error from the two steps of the surrogacy analysis. For comparison, the third column, for each gender, reports the actual treatment effect of the internship quasi-experiment. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 4: Family Formation Outcomes

	Single		Partnered	
	Males	Females	Males	Females
	(1)	(2)	(3)	(4)
A. Partnership				
Average Treatment Effect	-0.0025	0.0142	0.0153	-0.0078
	(0.0242)	(0.0231)	(0.0145)	(0.0129)
Constant	0.8003***	0.7576***	0.9244***	0.9215***
	(0.0168)	(0.0163)	(0.0114)	(0.0093)
Individuals	917	1,227	840	1,472
B. Number of Children				
Average Treatment Effect	-0.0001	0.1422***	-0.0376	-0.0142
	(0.0630)	(0.0549)	(0.0642)	(0.0464)
Constant	1.1685***	1.2374***	2.0891***	2.1001***
	(0.0449)	(0.0387)	(0.0492)	(0.0324)
Individuals	919	1,229	844	1,473
C. One Child or More				
Average Treatment Effect	0.0202	0.0464*	0.0221	-0.0068
	(0.0289)	(0.0244)	(0.0172)	(0.0126)
Constant	0.6567***	0.6961***	0.9138***	0.9322***
	(0.0208)	(0.0177)	(0.0133)	(0.0088)
Individuals	919	1,229	844	1,473
D. More than One Child				
Average Treatment Effect	-0.0165	0.0709***	0.0093	-0.0190
	(0.0290)	(0.0260)	(0.0254)	(0.0189)
Constant	0.4262***	0.4524***	0.7869***	0.8244***
	(0.0206)	(0.0185)	(0.0188)	(0.0131)
Individuals	919	1,229	844	1,473

Notes: This table studies the long-run effects of the lottery on family formation choices based on equation (2). We split the sample by whether individuals were single or partnered in the pre-period and study the probability of becoming partnered, the number of children, the probability of having one child or more, and the probability of having more than one child. To reduce potential measurement error in the partnership outcome, we make the adjustment that, if an individual has a missing value for the partner's identification number in a given period but the two individuals are reported as partners in both adjacent periods (before and after), we assign them as partners in that period as well. Robust standard errors, clustered at the individual level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 5: Marriage Market Matching

A. Matching Predictions

	Female Physicians (1)	Male Physicians (2)
Average ratio of the probability a potential partner matches with a physician in treatment group relative to control group	1.25***	1.03***
	(0.0002)	(0.0001)
Observations	8,880,950	8,348,438

Panel B: Partnership Characteristics

	Males	Females
Age Gap of Husband relative to Wife		
Average Treatment Effect	-0.1673	0.6175**
	(0.2478)	(0.2638)
Constant	-1.5456***	1.5100***
	(0.1743)	(0.1847)
Individuals	777	1,011
Assortative Mating by Medical Degree		
Average Treatment Effect	-0.0012	-0.0619**
	(0.0395)	(0.0314)
Constant	0.3483***	0.2502***
	(0.0279)	(0.0234)
Individuals	659	831

Notes: This table studies marriage market matching of individuals who were single at the baseline period. Panel A studies the potential effects on matching patterns among single physicians by constructing measures of matching likelihoods based on a set of observables for the pool of their potential partners. We first take, from the general population, an approximate pool of individuals who could be potential partners for the subjects of our quasi-experiment. For our single female physicians, we take the pool of all males of ages 30-50, and, for our single male physicians, we take the pool of all females of ages 25-45. We then predict, using lottery years 6-10, the match probability for each person i in the partner pool of marrying a person in our subject pool. We split the single subject pool on two dimensions: whether the physician belongs to the treatment group or the control group (indexed by $l \in \{t, c\}$) and whether the physician is male or female (indexed by $n \in \{m, f\}$). We then let $D_{i(l,n)}$ denote an indicator for a potential partner i marrying a physician from group (l, n), and we estimate a set of logistic regressions: $\Pr(D_{i(l,n)} = 1|X_i) = F(\beta_{(l,n)}X_i)$, where $F(z) = \exp(z)/[1 + \exp(z)]$. In the observables vector, X_i, we include a third-order polynomial in age and whether the potential partner also holds a medical degree, and we additionally include year fixed effects. Finally, we calculate treatment/control ratios of the predictions: $\Pr(D_{i(t,n)} = 1|X_i)/\Pr(D_{i(c,n)} = 1|X_i)$, for male/female subjects ($n \in \{m, f\}$). Panel A displays these ratios. Ratios that meaningfully deviate from one represent evidence of selection on observables, which implies differential matching in the marriage market as a result of the quasi-experiment. Panel B analyzes as outcomes the characteristics of the actual partners with whom our single subjects match using equation (2), and it includes only observations with non-missing partners in a given period. Robust standard errors, clustered at the individual level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 6: Mechanisms—Family vs. Career Tradeoff

A. Human Capital and Fertility Interactions among Pre-Lottery Singles

	Males (1)	Females (2)
Average Treatment Effect	-0.0058	0.0702**
	(0.0308)	(0.0285)
Constant	0.3298***	0.4087***
	(0.0221)	(0.0200)
Individuals	775	1,013

B. Proximity to Grandmothers

	Distance (Kilometers)	
	Males (1)	Females (2)
Average Treatment Effect	1.70	-7.46**
	(4.62)	(3.7168)
Constant	75.36***	77.81***
	(3.28)	(2.77)
Individuals	1,603	2,397

	Close Proximity (below mean)		Close Proximity & High Fertility		Close Proximity & Low Fertility	
	Males (1)	Females (2)	Males (3)	Females (4)	Males (5)	Females (6)
Average Treatment Effect	-0.0306	0.0505**	-0.0078	0.0466**	-0.0198	0.0054
	(0.0244)	(0.0200)	(0.0240)	(0.0204)	(0.0201)	(0.0159)
Constant	0.6611***	0.6316***	0.3849***	0.4135***	0.2554***	0.2041***
	(0.0172)	(0.0145)	(0.0173)	(0.0145)	(0.0148)	(0.0114)
Individuals	1,603	2,397	1,538	2,332	1,538	2,332

C. Search Behavior and Mobility

	Commuting Distance (Kilometers)		Propensity to Migrate across Labor Markets (Pre-Lottery Singles)	
	Males (1)	Females (2)	Males (3)	Females (4)
Average Treatment	2.5086**	-0.5712	-0.0641**	-0.0128
	(1.2160)	(0.9888)	(0.0304)	(0.0266)
Constant	26.5511***	26.5724***	0.4671***	0.4400***
	(0.8325)	(0.7500)	(0.0216)	(0.0187)
Individuals	1,705	2,536	959	1,277

Notes: This table provides investigations of the hypothesis that men and women differentially tackle the tension between family and career in their responses to initial sorting. In panel A, we study human capital and fertility interactions among single women for whom we found effects on fertility. We test whether it is indeed the case that the women who have more children also invest less in human capital. We analyze, as an outcome, the joint probability of not earning a PhD and having more than one child. In panel B, we study proximity to grandmothers. We first study as an outcome the distance to the physician's mother. To assess whether these patterns could indeed link to family considerations, we then study as an outcome the joint probability of living close to one's mother (based on a dummy for distance being below the sample mean) and having higher fertility (based on our measure of having more than one child that we found to be the active fertility margin). In panel C, we study outcomes that reflect search behavior and mobility. We first analyze the effect of the lottery on commuting distance. We then investigate internal migration across labor markets. To this end, we consider the sample of pre-lottery singles who face the decision of forming a new household. We then study the physician's propensity to reside, in the long run, within the pre-lottery location of their new spouse. All regressions use the specification of equation (2). Robust standard errors, clustered at the individual level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 7: Mechanisms—Workplace and Mentorship

A. "Excess Sensitivity" to Workplace

Degree of Exposure to Employer Intensity	Males (1)	Females (2)	*Sensitivity to Employer Intensity*	Males (3)	Females (4)
Treatment Effect	-0.1673***	-0.1445***	Employer Gender-Specific	0.4391***	0.5834***
	(0.0081)	(0.0057)	Placement	(0.0526)	(0.0408)
Constant	0.4709***	0.4461***	Constant	0.2099***	0.1550***
	(0.0056)	(0.0041)		(0.0226)	(0.0168)
Individuals	1,779	3,097	Individuals	2,484	4,260

B. Mentorship: Exposure to Female Mentors and Evaluations

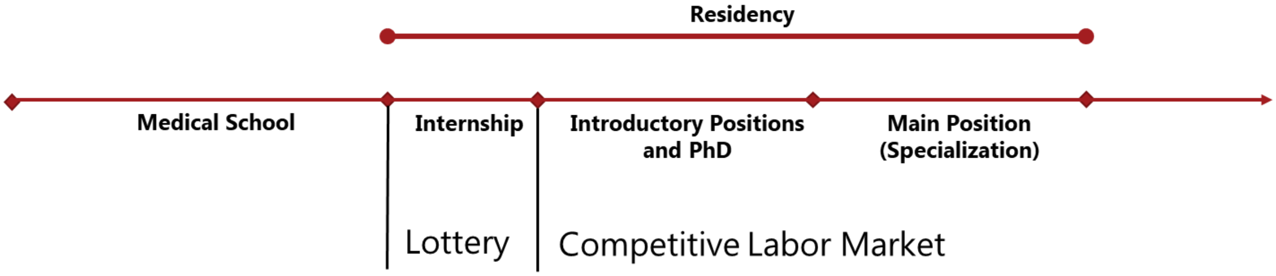
	Probability of Female Mentor		*Probability of Female Head of Educational Program*		*Evaluation of Mentorship*	
	Males (1)	Females (2)	Males (3)	Females (4)	Males (5)	Females (6)
Treatment Effect	-0.1017***	-0.1189***	-0.1043***	-0.0828***	-0.1509	-0.4264***
	(0.0286)	(0.0219)	(0.0281)	(0.0216)	(0.1019)	(0.0807)
Constant	0.4376***	0.4976***	0.4090***	0.4262***	7.0663***	7.1574***
	(0.0188)	(0.0154)	(0.0185)	(0.0152)	(0.0672)	(0.0566)
Individuals	1,177	2,016	1,177	2,016	1,177	2,016
Standard Deviation (SD)					1.75	1.68
Effect/SD					-0.09	-0.25

Notes: This table investigates the role of employer-side factors as an explanation for the gender divergence we uncover. Panel A studies sensitivity to employers by gender. Employer intensity is defined as a leave-one-out mean of the hospital departments' propensity to place their interns in a university hospital in their subsequent positions. We provide estimates for interns' exposure to employer intensity (columns 1-2), and we then provide estimates for the interns' sensitivity to employer intensity by gender by regressing one's own probability of being employed at a university hospital in their next position on the intensity of their internship employer (columns 3-4). Panel B investigates the potential role of a particular workplace characteristic, that is, the mentorship it provides. Using the internship exit surveys, we study the probability of being assigned a female mentor, the probability that the head of the educational program is female, and the interns' evaluation of the mentorship they have received. Robust standard errors, clustered at the individual level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Online Appendix

Appendix A: Danish Physicians' Postgraduate Training

Appendix Figure A.1: Timeline



Notes: This figure summarizes the timeline of Danish physicians' training, which captures the early stages of their careers.

Appendix B: Lottery Verification and Summary Statistics

Appendix Table B.1: Verification of Lottery

	Overall Sample (1)	Males (2)	Females (3)
Gender	0.0074		
	(0.0060)		
Age	0.0004	-0.0008	0.0014
	(0.0013)	(0.0020)	(0.0018)
Partnered	0.0086	0.0084	0.0089
	(0.0063)	(0.0100)	(0.0081)
Number of Children	-0.0030	-0.0039	-0.0033
	(0.0058)	(0.0099)	(0.0073)
GPA Rank	0.0048	0.0025	0.0068
	(0.0104)	(0.0162)	(0.0136)
Observations	10,017	3,939	6,078
R-Squared	0.0004	0.0003	0.0003
F-Statistic	0.74	0.25	0.48
p-Value	0.5959	0.9082	0.7507

Notes: This table tests the validity of the lottery in terms of random assignment. We run specifications that regress the graduating physicians' lottery rank on baseline characteristics available in our data. These include gender, age, an indicator for having a registered partner, number of children in the household, and high school GPA rank. Robust standard errors are reported in parentheses, and we also report the *p*-value of the *F*-test for the joint predictive power of the specifications we run. * $p < .10$, ** $p < .05$, *** $p < .01$

Appendix Table B.2: Analysis Sample Summary Statistics

	Control (1)	Treatment (2)	Difference (3)	*p*-value (4)
A. Overall Sample				
Female	0.5999	0.6114	-0.0115	0.3576
Partnered	0.4964	0.5079	-0.0115	0.3700
Age	28.5096	28.5206	-0.0111	0.8606
GPA Rank	0.5021	0.5047	-0.0026	0.7246
Number of Children	0.2669	0.2644	0.0025	0.8694
Number of Individuals	3,024	3,052		
B. Males				
Partnered	0.4636	0.4696	-0.0060	0.7681
Age	28.6455	28.5995	0.0460	0.6665
GPA Rank	0.5052	0.4986	0.0066	0.5871
Number of Children	0.2280	0.2184	0.0096	0.6654
Number of Individuals	1,210	1,186		
C. Females				
Partnered	0.5182	0.5322	-0.0140	0.3964
Age	28.4190	28.4705	-0.0516	0.5047
GPA Rank	0.5000	0.5085	-0.0086	0.3652
Number of Children	0.2928	0.2935	-0.0008	0.9682
Number of Individuals	1,814	1,866		

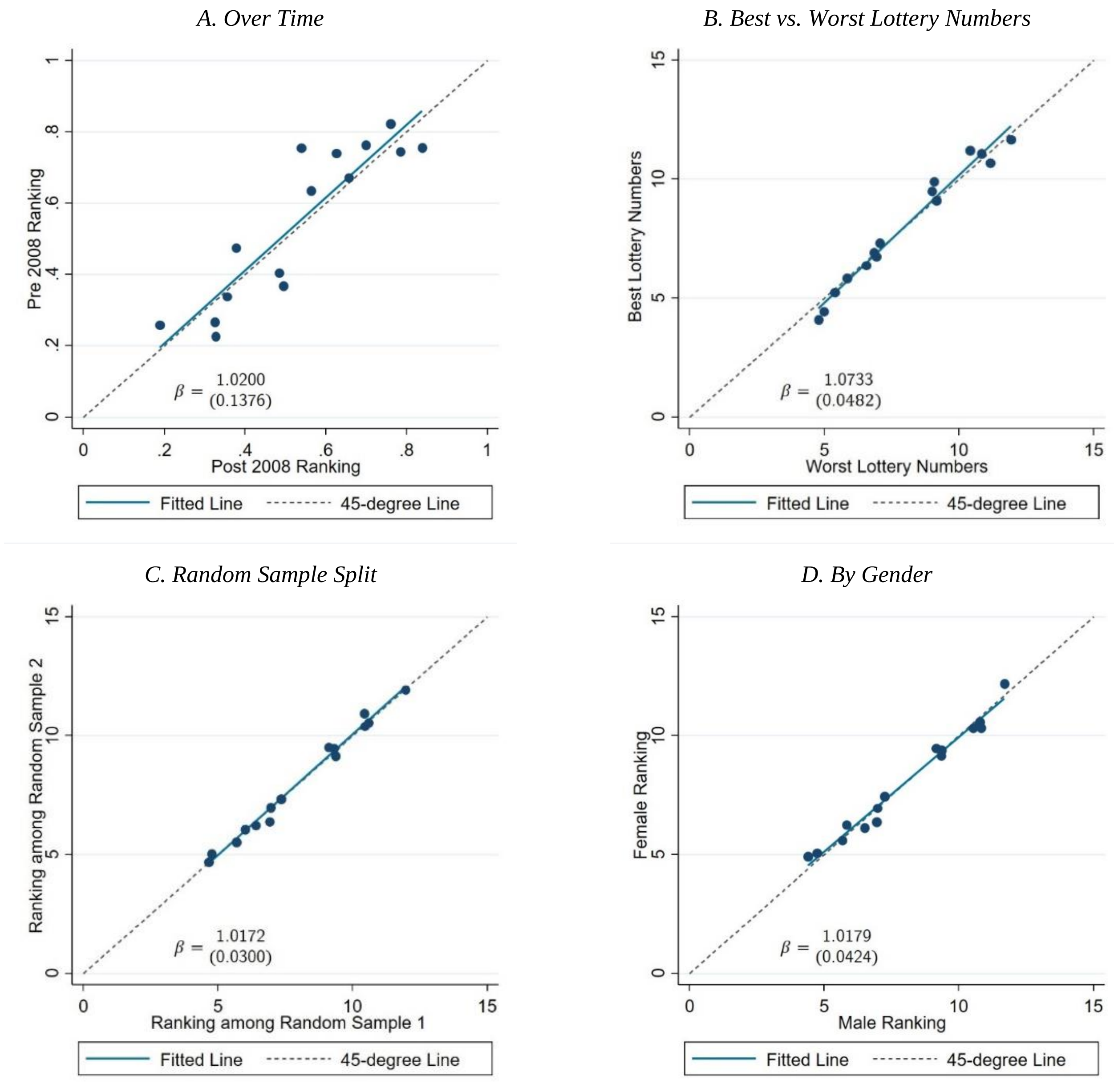
Notes: This table provides summary statistics for the analysis sample in the year prior to the internship lottery. Panel A provides statistics for the entire sample, and panels B and C split the sample by gender. Characteristics include gender, age, an indicator for having a registered partner (in cohabitation or marriage), number of children in the household, and high school GPA rank. Column 1 displays means for our control group, and column 2 displays means for our treatment group. Column 3 provides the differences between column 1 and column 2. Column 4 reports the *p*-values of the test statistics (*t*-statistics for continuous variables and *z*-statistics for binary variables) of the differences in column 3.

Appendix C: Labor Market Rankings

Appendix C.1: Labor Market Rankings for Random Sample Split

Local labor markets and the average characteristics of the jobs they offer have aspects that people may agree upon ("vertical" quality, e.g., interning in a teaching hospital) and aspects that could be individual specific ("horizontal" quality whose valuation can differ across individuals, e.g., a county's proximity to family). To investigate the degree to which the rankings of the labor markets are agreed upon among the new physicians (as compared to diverging across them due to individual specific preferences), we compare the rankings of labor markets across a random split of our analysis sample. If students tend to agree on the value of characteristics of labor markets, we would expect the overall average rankings of the two random subsamples to align on the 45-degree line, and, if preferences are completely idiosyncratic (an extreme case), there should be no systematic relationship across the two groups' rankings. Panel C of Appendix Figure C.1 shows that the average rankings of the local labor markets across the two groups line up around the 45-degree line, and we cannot reject the benchmark null of a coefficient of one, which represents ranking comparability. We note that while this finding suggests there is a degree of general agreement over labor market rankings across students, it does not mean there are no components of idiosyncratic preferences (over "horizontal" quality). In fact, the observation that the two groups' rankings do not perfectly align on the 45-degree is, in itself, an indication of the natural presence of individual-specific considerations.

Appendix Figure C.1: Labor Market Rankings

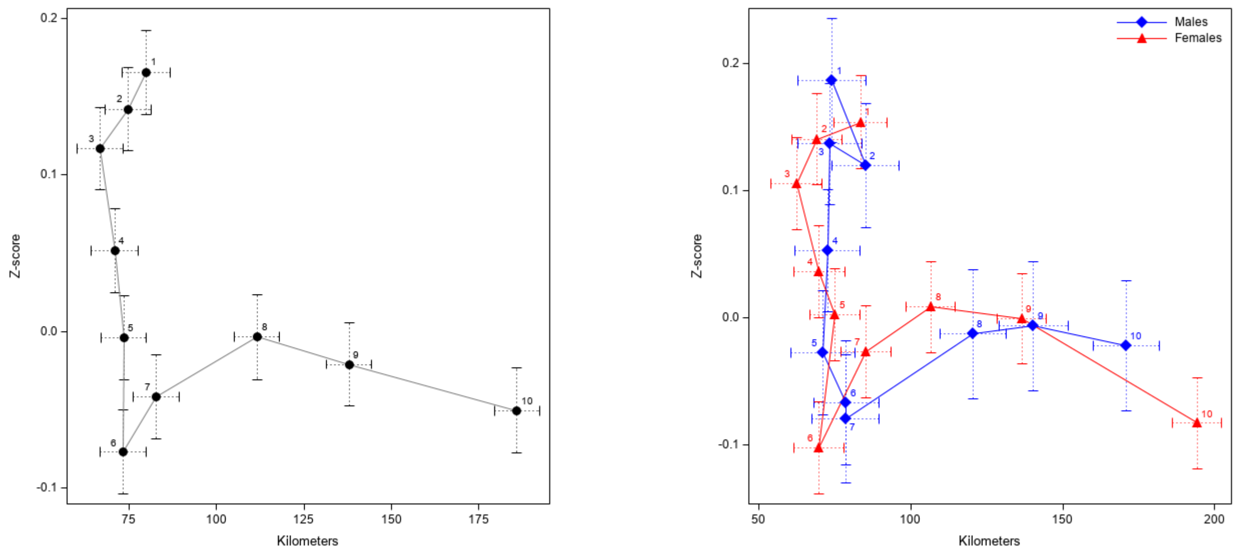


Notes: This figure makes several comparisons of the effective rankings of local labor markets. In panel A, location-based preferences, as revealed through choices, are constructed such that we characterize the desirability of a labor market (i.e., a county) based on the average lottery rank of the interns who choose to sort into it. Panel A compares the average rankings across earlier cohorts and later cohorts. In panels B-D, we use the information we have for earlier cohorts about students' binding pre-placement rankings of all local labor markets, as reported in priority lists. Panel B compares the average rankings of those with the best lottery numbers (the bottom 30 percent) with the average rankings of those with the worst lottery numbers (the top 30 percent). Panel C compares the average rankings of labor markets across a random split of our analysis sample. Panel D compares females' and males' priority rankings over entry-level local labor markets (where the counterpart that uses the market desirability measure appears in panel A of Figure 5). We assign each local labor market its average priority by gender, and we then compare these priority rankings across males and females. In all panels, each dot represents a local labor market. We plot the fitted line, as well as the 45-degree line, which is the benchmark under non-differential rankings by gender. We also report the slope of the fitted line, where the benchmark of non-differential ranking is one.

Appendix D: Internship Period First Stage

Appendix Figure D.1: Distance vs. Quality Tradeoff

A. Overall Sample *B. Split by Gender*



Notes: This figure replicates panels E-F of Figure 1, but we group subjects into ten equal-sized bins based on their lottery ranks. Each dot represents a decile (whose number is displayed in the figure), and it plots the average values within that decile for the internship characteristics of relocation distance (on the x-axis) and a z-score of quality (on the y-axis), along with their corresponding 95-percent confidence intervals.

Appendix D.1: Unpacking the Treatment Bundle

In Section 5.1, we describe how we can shed light on the relative role of different dimensions of the treatment by leveraging comparisons to the middle experimental group of intermediate lottery ranks. This appendix formalizes the intuitive description we provide in the text. A simple way to think of this setting is with the same basic logic as a traditional difference-in-differences setting, as follows. Let us split internships dichotomously on the two dimensions we consider (say, based on their mean values): into internships whose ranked quality (q) is high (1) or low (0) and internships whose distance from the origin (d) is far (1) or close (0). Assume that a long run outcome y_i is determined by these two dimensions, so that:

$$y_i = \beta^q \times \mathbb{I}(q = 0)_i + \beta^d \times \mathbb{I}(d = 1)_i + \epsilon_{it}.$$

For simplicity, further assume that, for individuals in the *control* group ($i \in C$) we have $q = 1$ and $d = 0$, for individuals in the *middle* group ($i \in M$) we have $q = 0$ and $d = 0$, and for individuals in the *treatment* group ($i \in T$) we have $q = 0$ and $d = 1$ (whereas all of these could be made probabilistic in a straightforward way). This structure assumes: (i) additivity, i.e., in practice, there are no economically meaningful complementarities across the two dimensions; (ii) exclusion, i.e., in practice, the composites of ranked quality and distance capture the bulk of the variation relevant for the long run outcomes (or are highly correlated with it). Under these assumptions, this analysis offers a complete decomposition of the total effect, whereas the decomposition would be only partial if these "identifying" assumptions are meaningfully violated.

With this structure, the total effect will be identified by a comparison between the treatment group and the control group: $E(y_i | i \in T) - E(y_i | i \in C) = \beta^q + \beta^d$; the first difference between the middle group and the control group would identify the portion attributed to ranked quality: $E(y_i | i \in M) - E(y_i | i \in C) = \beta^q$; and the second difference between the treatment group and the middle group would identify the portion attributed to geographic markets: $E(y_i | i \in T) - E(y_i | i \in M) = \beta^d$.[1]

[1] In the presence of meaningful interaction terms across dimensions in longer run effects, $E(y_i | i \in M) - E(y_i | i \in C)$ would capture the impact of ranked quality within urban markets, and $E(y_i | i \in T) - E(y_i | i \in M)$ would capture the effect of geographic markets within internships of low-ranked quality.

Appendix E:

Alternative Specifications

Appendix Table E.1: Research Design—Alternative Specifications

A. Sorting into Less Desirable Local Labor Markets

All

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	0.0527***	0.0586***	0.0538***	0.0469***	0.0476***	0.0507***	0.0773***
	(0.0149)	(0.0135)	(0.0122)	(0.0114)	(0.0106)	(0.0117)	(0.0164)
Middle	0.0180	0.0076	0.0181	0.0120	0.0165	0.0130	
	(0.0117)	(0.0112)	(0.0111)	(0.0115)	(0.0128)	(0.0112)	
Constant	0.1710***	0.1737***	0.1689***	0.1723***	0.1699***	0.1711***	0.1536***
	(0.0100)	(0.0090)	(0.0082)	(0.0077)	(0.0071)	(0.0079)	(0.0091)
Individuals, incl. middle	7,037	7,037	7,037	7,037	7,037	7,037	7,037
Individuals, excl. middle	2,852	3,557	4,250	4,941	5,642	4,668	

Males

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	0.0107	0.0104	0.0152	0.0190	0.0236	0.0021	0.0290
	(0.0229)	(0.0209)	(0.0191)	(0.0179)	(0.0167)	(0.0181)	(0.0257)
Middle	0.0221	-0.0009	0.0069	0.0049	0.0104	0.0164	
	(0.0188)	(0.0182)	(0.0180)	(0.0184)	(0.0205)	(0.0185)	
Constant	0.1805***	0.1934***	0.1883***	0.1876***	0.1841***	0.1926***	0.1811***
	(0.0160)	(0.0147)	(0.0134)	(0.0125)	(0.0115)	(0.0162)	(0.0146)
Individuals, incl. middle	2,798	2,798	2,798	2,798	2,798	2,798	2,798
Individuals, excl. middle	1,138	1,436	1,706	1,948	2,230	1,842	

Females

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	0.0812***	0.0918***	0.0802***	0.0653***	0.0634***	0.0734***	0.1096***
	(0.0196)	(0.0176)	(0.0159)	(0.0147)	(0.0136)	(0.0151)	(0.0213)
Middle	0.0155	0.0137	0.0259*	0.0165	0.0203	0.0200	
	(0.0149)	(0.0141)	(0.0140)	(0.0147)	(0.0164)	(0.0143)	
Constant	0.1645***	0.1602***	0.1558***	0.1623***	0.1606***	0.1590***	0.1352***
	(0.0127)	(0.0113)	(0.0103)	(0.0097)	(0.0090)	(0.0099)	(0.0116)
Individuals, incl. middle	4,239	4,239	4,239	4,239	4,239	4,239	4,239
Individuals, excl. middle	1,714	2,121	2,544	2,993	3,412	2,826	

B. Human Capital Investment—Obtaining a Medical PhD

All

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	-0.0222*	-0.0186*	-0.0147	-0.0083	-0.0043	-0.0142	-0.0183
	(0.0117)	(0.0103)	(0.0096)	(0.0089)	(0.0083)	(0.0092)	(0.0127)
Middle	-0.0077	0.0021	-0.0038	-0.0006	0.0013	-0.0042	
	(0.0099)	(0.0093)	(0.0091)	(0.0093)	(0.0104)	(0.0092)	
Constant	0.1390***	0.1337***	0.1359***	0.1331***	0.1314***	0.1361***	0.1391***
	(0.0086)	(0.0075)	(0.0069)	(0.0064)	(0.0059)	(0.0066)	(0.0075)
Individuals, incl. middle	6,386	6,386	6,386	6,386	6,386	6,386	6,386
Individuals, excl. middle	2,588	3,224	3,857	4,482	5,124	4,322	

Males

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	-0.0021	0.0105	0.0093	0.0174	0.0177	0.0136	0.0115
	(0.0219)	(0.0191)	(0.0177)	(0.0166)	(0.0153)	(0.0170)	(0.0236)
Middle	-0.0094	0.0125	0.0049	0.0034	0.0129	0.0052	
	(0.0179)	(0.0166)	(0.0165)	(0.0168)	(0.0189)	(0.0166)	
Constant	0.1819***	0.1670***	0.1711***	0.1687***	0.1661***	0.1695***	0.1701***
	(0.0155)	(0.0134)	(0.0124)	(0.0114)	(0.0106)	(0.0118)	(0.0136)
Individuals, incl. middle	2,538	2,538	2,538	2,538	2,538	2,538	2,538
Individuals, excl. middle	1,040	1,304	1,551	1,770	2,027	1,674	

Females

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	-0.0361***	-0.0397***	-0.0325***	-0.0262***	-0.0195**	-0.0335***	-0.0390***
	(0.0123)	(0.0110)	(0.0104)	(0.0097)	(0.0091)	(0.0100)	(0.0136)
Middle	-0.0056	-0.0035	-0.0085	-0.0046	-0.0077	-0.0112	
	(0.0111)	(0.0106)	(0.0104)	(0.0106)	(0.0117)	(0.0105)	
Constant	0.1095***	0.1106***	0.1121***	0.1095***	0.1083***	0.1139***	0.1185***
	(0.0096)	(0.0087)	(0.0080)	(0.0073)	(0.0067)	(0.0077)	(0.0084)
Individuals, incl. middle	3,848	3,848	3,848	3,848	3,848	3,848	3,848
Individuals, excl. middle	1,548	1,920	2,306	2,712	3,097	2,558	

C. Labor Market Position—Affiliation with a University Hospital

All

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	-0.0328**	-0.0396***	-0.0369***	-0.0381***	-0.0343***	-0.0392***	-0.0551***
	(0.0147)	(0.0131)	(0.0121)	(0.0112)	(0.0104)	(0.0115)	(0.0161)
Middle	-0.0083	-0.0004	-0.0037	0.0010	-0.0049	0.0006	
	(0.0120)	(0.0114)	(0.0113)	(0.0116)	(0.0129)	(0.0114)	
Constant	0.4509***	0.4496***	0.4520***	0.4524***	0.4542***	0.4522***	0.4669***
	(0.0104)	(0.0093)	(0.0085)	(0.0079)	(0.0074)	(0.0082)	(0.0093)
Individuals, incl. middle	7,616	7,616	7,616	7,616	7,616	7,616	7,616
Individuals, excl. middle	3,085	3,850	4,601	5,346	6,107	5,054	

Males

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	-0.0130	-0.0028	-0.0063	-0.0104	-0.0135	-0.0127	-0.0237
	(0.0234)	(0.0207)	(0.0190)	(0.0177)	(0.0166)	(0.0183)	(0.0256)
Middle	-0.0273	-0.0082	-0.0092	-0.0084	-0.0155	-0.0144	
	(0.0190)	(0.0181)	(0.0180)	(0.0186)	(0.0206)	(0.0182)	
Constant	0.4760***	0.4618***	0.4627***	0.4633***	0.4657***	0.4663***	0.4690***
	(0.0164)	(0.0146)	(0.0134)	(0.0125)	(0.0117)	(0.0129)	(0.0148)
Individuals, incl. middle	3,017	3,017	3,017	3,017	3,017	3,017	3,017
Individuals, excl. middle	1,223	1,540	1,830	2,090	2,401	1,979	

Females

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	-0.0459**	-0.0646***	-0.0577***	-0.0562***	-0.0480***	-0.0565***	-0.0758***
	(0.0188)	(0.0169)	(0.0155)	(0.0143)	(0.0134)	(0.0148)	(0.0206)
Middle	0.0046	0.0050	0.0001	0.0071	0.0021	0.0105	
	(0.0155)	(0.0148)	(0.0145)	(0.0149)	(0.0165)	(0.0147)	
Constant	0.4340***	0.4413***	0.4448***	0.4453***	0.4466***	0.4429***	0.4655***
	(0.0134)	(0.0121)	(0.0111)	(0.0103)	(0.0096)	(0.0106)	(0.0120)
Individuals, incl. middle	4,599	4,599	4,599	4,599	4,599	4,599	4,599
Individuals, excl. middle	1,862	2,310	2,771	3,256	3,706	3,075	

D. Occupational Choice—Gender-Represented Specialty

All

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	0.0224***	0.0222***	0.0193***	0.0182***	0.0172***	0.0198***	0.0274***
	(0.0072)	(0.0066)	(0.0060)	(0.0055)	(0.0052)	(0.0057)	(0.0080)
Middle	0.0149***	0.0073	0.0075	0.0064	0.0074	0.0078	
	(0.0055)	(0.0053)	(0.0053)	(0.0055)	(0.0061)	(0.0054)	
Constant	0.0682***	0.0724***	0.0728***	0.0733***	0.0732***	0.0724***	0.0679***
	(0.0046)	(0.0042)	(0.0039)	(0.0036)	(0.0034)	(0.0037)	(0.0043)
Individuals, incl. middle	7,037	7,037	7,037	7,037	7,037	7,037	7,037
Individuals, excl. middle	2,852	3,557	4,250	4,941	5,642	4,668	

Males

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	0.0093	0.0059	0.0031	0.0117	0.0102	0.0099	0.0103
	(0.0106)	(0.0098)	(0.0086)	(0.0086)	(0.0080)	(0.0088)	(0.0121)
Middle	0.0193**	0.0085	0.0065	0.0084	0.0080	0.0080	
	(0.0087)	(0.0086)	(0.0092)	(0.0089)	(0.0099)	(0.0087)	
Constant	0.0606***	0.0682***	0.0706***	0.0672***	0.0682***	0.0679***	0.0687***
	(0.0072)	(0.0068)	(0.0063)	(0.0058)	(0.0055)	(0.0059)	(0.0069)
Individuals, incl. middle	2,798	2,798	2,798	2,798	2,798	2,798	2,798
Individuals, excl. middle	1,138	1,436	1,706	1,948	2,230	1,842	

Females

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	0.0312***	0.0336***	0.0282***	0.0226***	0.0219***	0.0264***	0.0388***
	(0.0097)	(0.0088)	(0.0079)	(0.0072)	(0.0067)	(0.0075)	(0.0105)
Middle	0.0119*	0.0064	0.0102	0.0052	0.0071	0.0078	
	(0.0071)	(0.0067)	(0.0067)	(0.0070)	(0.0078)	(0.0068)	
Constant	0.0734***	0.0752***	0.0743***	0.0773***	0.0766***	0.0754***	0.0673***
	(0.0060)	(0.0054)	(0.0050)	(0.0046)	(0.0043)	(0.0047)	(0.0056)
Individuals, incl. middle	4,239	4,239	4,239	4,239	4,239	4,239	4,239
Individuals, excl. middle	1,714	2,121	2,544	2,993	3,412	2,826	

E. Probability of Having a Partner among Pre-Lottery Singles

All

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	0.0112	0.0140	0.0070	0.0046	0.0019	-0.0012	0.0018
	(0.0203)	(0.0183)	(0.0168)	(0.0157)	(0.0146)	(0.0163)	(0.0223)
Middle	0.0007	0.0031	-0.0006	0.0135	0.0010	0.0128	
	(0.0167)	(0.0160)	(0.0158)	(0.0162)	(0.0179)	(0.0158)	
Constant	0.7751***	0.7728***	0.7760***	0.7721***	0.7769***	0.7738***	0.7769***
	(0.0144)	(0.0130)	(0.0118)	(0.0110)	(0.0103)	(0.0113)	(0.0128)
Individuals, incl. middle	3,574	3,574	3,574	3,574	3,574	3,574	3,574
Individuals, excl. middle	1,469	1,814	2,144	2,481	2,858	2,337	

Males

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	0.0033	0.0105	-0.0025	-0.0143	-0.0140	-0.0148	-0.0172
	(0.0301)	(0.0265)	(0.0242)	(0.0229)	(0.0215)	(0.0236)	(0.0330)
Middle	-0.0147	-0.0120	-0.0240	-0.0221	-0.0191	-0.0140	
	(0.0245)	(0.0235)	(0.0234)	(0.0243)	(0.0275)	(0.0236)	
Constant	0.7983***	0.7933***	0.8003***	0.8019***	0.7995***	0.7999***	0.7989***
	(0.0209)	(0.0186)	(0.0168)	(0.0157)	(0.0148)	(0.0162)	(0.0187)
Individuals, incl. middle	1,505	1,505	1,505	1,505	1,505	1,505	1,505
Individuals, excl. middle	608	774	917	1,047	1,216	991	

Females

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	0.0172	0.0165	0.0142	0.0183	0.0139	0.0088	0.0160
	(0.0274)	(0.0251)	(0.0231)	(0.0215)	(0.0199)	(0.0222)	(0.0301)
Middle	0.0121	0.0148	0.0172	0.0394*	0.0162	0.0326	
	(0.0227)	(0.0218)	(0.0213)	(0.0216)	(0.0236)	(0.0213)	
Constant	0.7581***	0.7572***	0.7576***	0.7503***	0.7599***	0.7545***	0.7608***
	(0.0196)	(0.0179)	(0.0163)	(0.0152)	(0.0141)	(0.0156)	(0.0174)
Individuals, incl. middle	2,069	2,069	2,069	2,069	2,069	2,069	2,069
Individuals, excl. middle	861	1,040	1,227	1,434	1,642	1,346	

F. Probability of Having More than One Child among Pre-Lottery Singles

All

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	0.0415*	0.0383*	0.0331*	0.0174	0.0122	0.0234	0.0310
	(0.0235)	(0.0212)	(0.0194)	(0.0181)	(0.0168)	(0.0186)	(0.0258)
Middle	0.0155	0.0212	0.0229	0.0263	0.0138	0.0198	
	(0.0193)	(0.0184)	(0.0181)	(0.0186)	(0.0205)	(0.0183)	
Constant	0.4425***	0.4400***	0.4411***	0.4461***	0.4526***	0.4457***	0.4447***
	(0.0166)	(0.0150)	(0.0138)	(0.0128)	(0.0119)	(0.0132)	(0.0149)
Individuals, incl. middle	3,581	3,581	3,581	3,581	3,581	3,581	3,581
Individuals, excl. middle	1,471	1,816	2,148	2,486	2,864	2,341	

Males

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	0.0086	-0.0050	-0.0165	-0.0298	-0.0373	-0.0250	-0.0328
	(0.0357)	(0.0318)	(0.0290)	(0.0272)	(0.0253)	(0.0279)	(0.0388)
Middle	-0.0102	-0.0070	-0.0000	0.0048	0.0033	0.0047	
	(0.0291)	(0.0279)	(0.0276)	(0.0285)	(0.0318)	(0.0280)	
Constant	0.4254***	0.4259***	0.4262***	0.4302***	0.4356***	0.4278***	0.4375***
	(0.0251)	(0.0226)	(0.0206)	(0.0195)	(0.0182)	(0.0199)	(0.0227)
Individuals, incl. middle	1,508	1,508	1,508	1,508	1,508	1,508	1,508
Individuals, excl. middle	609	775	919	1,050	1,219	993	

Females

	Percentile						
	20	*25*	*30*	*35*	*40*	Tercile (*33*)	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	0.0648**	0.0711**	0.0709***	0.0529**	0.0488**	0.0597**	0.0783**
	(0.0311)	(0.0282)	(0.0260)	(0.0239)	(0.0223)	(0.0248)	(0.0343)
Middle	0.0342	0.0404*	0.0376	0.0412*	0.0193	0.0297	
	(0.0256)	(0.0245)	(0.0240)	(0.0245)	(0.0268)	(0.0242)	
Constant	0.4551***	0.4506***	0.4524***	0.4579***	0.4653***	0.4590***	0.4496***
	(0.0222)	(0.0201)	(0.0185)	(0.0169)	(0.0158)	(0.0175)	(0.0198)
Individuals, incl. middle	2,073	2,073	2,073	2,073	2,073	2,073	2,073
Individuals, excl. middle	862	1,041	1,229	1,436	1,645	1,348	

Notes: These tables investigate the robustness of our design by studying the effects on our main outcomes when we vary the percentiles that define the treatment and control groups. Columns 1-5 report estimates for long-run effects, based on specification (2), for thresholds that vary in five percentage-point increments. Column 3 corresponds to our main specification. Column 6 also reports estimates where the treatment, control, and middle groups are split at the 33rd and 67th percentiles (as a potentially natural benchmark). Column 7 estimates a version of specification (2) that is linear in lottery rank. The two bottom rows in each estimation report sample sizes, depending on whether the estimation includes or excludes the middle group. Robust standard errors, clustered at the individual level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Appendix Table E.2: Effects of Early Career Choices on Longer Run Outcomes—Graduation Round Fixed Effects

A. Sorting into Less Desirable Local Labor Markets

	All (1)	Males (2)	Females (3)
Treat	0.0538***	0.0139	0.0801***
	(0.0122)	(0.0190)	(0.0159)
Constant	0.1689***	0.1890***	0.1558***
	(0.0081)	(0.0133)	(0.0102)
Individuals	4,250	1,706	2,544

B. Obtaining a Medical PhD

	All (1)	Males (2)	Females (3)
Treat	-0.0143	0.0101	-0.0312***
	(0.0096)	(0.0177)	(0.0103)
Constant	0.1357***	0.1706***	0.1114***
	(0.0069)	(0.0124)	(0.0079)
Individuals	3,857	1,551	2,306

C. Affiliation with a University Hospital

	All (1)	Males (2)	Females (3)
Treat	-0.0392***	-0.0147	-0.0549***
	(0.0113)	(0.0180)	(0.0145)
Constant	0.4531***	0.4669***	0.4434***
	(0.0079)	(0.0128)	(0.0101)
Individuals	4,601	1,830	2,771

D. Occupational Choice—Gender-Represented Specialty

	All (1)	Males (2)	Females (3)
Treat	0.0190***	0.0062	0.0264***
	(0.0059)	(0.0092)	(0.0077)
Constant	0.0730***	0.0708***	0.0752***
	(0.0039)	(0.0063)	(0.0049)
Individuals	4,250	1,706	2,544

E. Probability of Having a Partner among Pre-Lottery Singles

	All (1)	Males (2)	Females (3)
Treat	0.0068	-0.0105	0.0158
	(0.0167)	(0.0236)	(0.0229)
Constant	0.7761***	0.8043***	0.7568***
	(0.0117)	(0.0162)	(0.0162)
Individuals	2,144	917	1,227

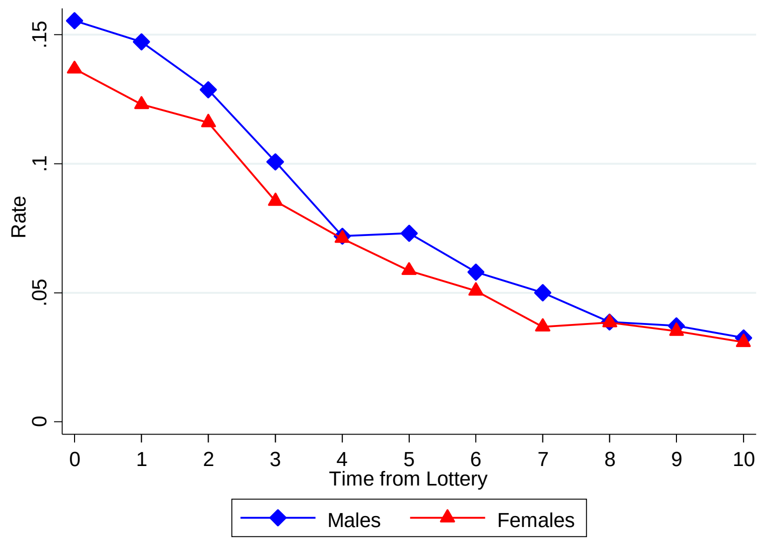
F. Probability of Having More than One Child among Pre-Lottery Singles

	All (1)	Males (2)	Females (3)
Treat	0.0321*	-0.0244	0.0708***
	(0.0193)	(0.0288)	(0.0257)
Constant	0.4416***	0.4302***	0.4524***
	(0.0136)	(0.0202)	(0.0183)
Individuals	2,148	919	1,229

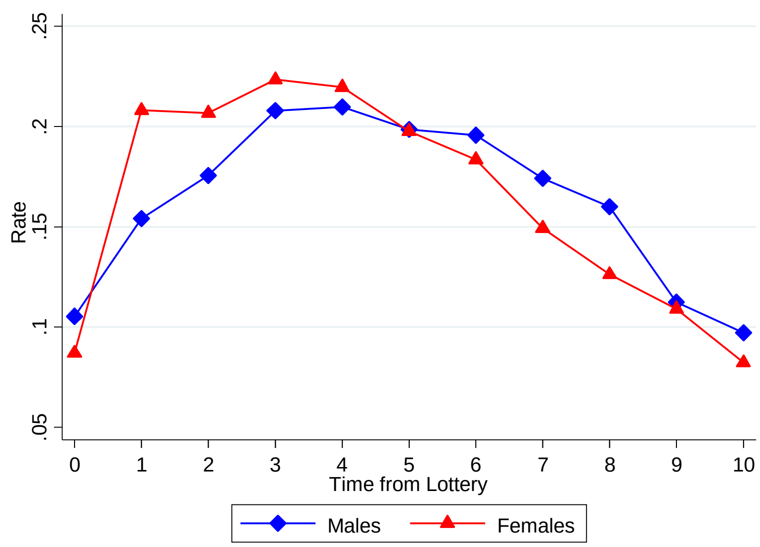
Notes: These tables investigate the robustness of the results for our main long-run outcomes to the inclusion of graduation round fixed effects based on specification (2). Robust standard errors, clustered at the individual level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Appendix Figure E.1: Baseline Dynamics in Marriage Market Choices

A. Household Formation—Change in Partnership Status

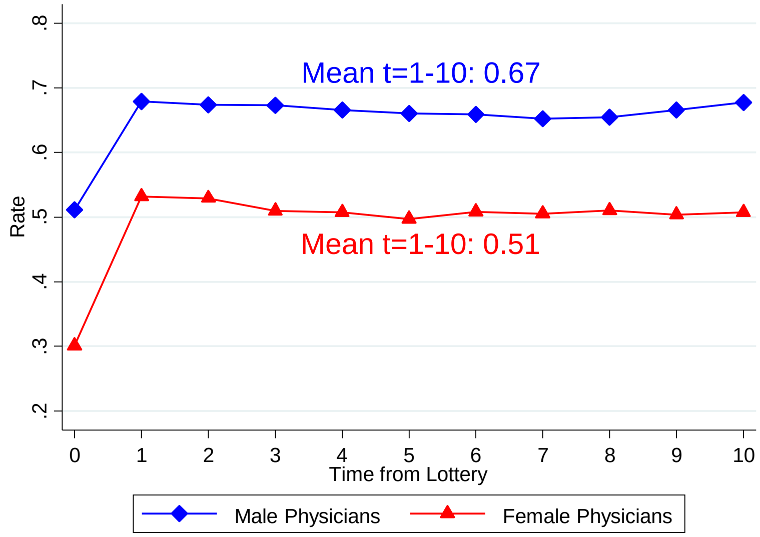


B. Fertility—Change in Number of Children



Notes: These figures illustrate how partnership rates and fertility evolve among physicians after graduation from medical school. Panel A displays the change in partnership status. It plots averages of a dummy variable that assumes the value of 1 if a physician's partnership status changed from the previous year (in either the direction of starting or ending a registered partnership), and it assumes the value of 0 otherwise. Panel B displays the change in number of children across consecutive years.

Appendix Figure E.2: Family vs. Career—Baseline Dynamics in Physicians' Labor Market Specialization



Notes: This figure plots, among physicians who are partnered in a given period, the relative earnings of the physicians out of their households' overall earnings as a proxy for labor market vs. household production specialization.

Appendix F: Specialty Grouping

Appendix Table F.1

Specialty	Specialty Group
Panel A: Male-Represented	
Thorax Surgery	Surgery
Orthopedic Surgery	Surgery
General Surgery	Surgery
Neurosurgery	Surgery
Internal Medicine	Internal medicine
Clinical Biochemistry	Transverse specialties
Otorhinolaryngology	Surgery
Internal Medicine: Cardiology	Internal medicine
Ophthalmology	Surgery
Vascular Surgery	Surgery
Anesthesiology	Transverse specialties
Internal Medicine: Gastroenterology and Hepatology	Internal medicine
Urology	Surgery
Panel B: Female-Represented	
Internal Medicine: Hematology	Internal medicine
Clinical Microbiology	Transverse specialties
Neuro Medicine	Other
Clinical Immunology	Transverse specialties
Clinical Physiology and Nuclear Medicine	Transverse specialties
Occupational Medicine	Other
General Medicine	General medicine
Internal Medicine: Rheumatology	Internal medicine
Internal Medicine: Pulmonary Diseases	Internal medicine
Radiology	Transverse specialties
Internal Medicine: Endocrinology	Internal medicine
Plastic Surgery	Surgery
Psychiatry	Psychiatry
Internal Medicine: Nephrology	Internal medicine
Dermato-Venerology	Other
Clinical Pharmacology	Transverse specialties
Internal Medicine: Infectious Diseases	Internal medicine
Gynecology and Obstetrics	Surgery
Pathological Anatomy and Cytology	Transverse specialties
Public Medicine	Other
Pediatrics	Other
Clinical Oncology	Other
Internal Medicine: Geriatrics	Internal medicine
Forensic medicine	Other
Clinical Genetics	Transverse specialties
Child and Youth Psychiatry	Psychiatry

Notes: This table classifies medical specialties by gender representativeness based on the share of females within a specialty relative to their overall proportion. “Female-represented specialties” are specialties with a female share that is higher than this proportion, and “male-represented specialties” are specialties with a female share that is lower than this proportion.

Appendix G: Exit Surveys

Appendix G.1: Exit Surveys—Details

This appendix provides background information on the exit surveys. The questions in the surveys are grouped into seven overall categories. The survey questions changed in 2016, but the seven categories remained similar. The average responses for each of the seven categories for each hospital department are reported on the public website www.evaluer.dk, and they are available for students to obtain information on the quality of their future workplaces.

Appendix Tables G.1 and G.2 show the groupings of the individual questions from the old and new questionnaires into the seven overall categories. The individual questions are provided in Appendix Tables G.3-G.6 in Danish (original) and English (translated). To provide numerical scoring of a department, interns also report the names of their supervisors: the assigned mentor and the head of the educational program. We use these names to deduct the gender of the supervisors. To do so, we construct an algorithm based on first names, which works as follows. We construct a gender probability using the first names of all doctors in the authorization register, which includes their names and gender. A first name is defined as "male" if more than 70 percent of the individuals with the given first name are males, and, accordingly, a first name is defined as "female" if less than 30 percent of the individuals with the given first name are males. We extract the first names of the supervisors from the exit surveys and match their first names to the gender proxy constructed from the authorization register.

Appendix Table G.1: Evaluation Categories in Evaluations until 2015

Group	English (translated)	Danish (original)	Questions
1	Introduction	Introduktion	1-2
2	Supervision	Uddannelsesprogram	3-6
3	Daily guidance	Vejleder (Praksistutor)	7-11
4	Work organization	Arbejdstilrettelæggelse	12-17
5	Education	Øvrige forhold	18-22
6	Education	Samlet vurdering	23
7	Overall Assessment	Samlet vurdering	24

Notes: The evaluation scales range from 1 to 9. The individual questions are reported in Appendix Tables G.3 and G.4.

Appendix Table G.2: Evaluation Categories in Evaluations from 2016

Group	English (translated)	Danish (original)	Questions
1	Introduction	Introduktion	1-3
2	Supervision	Uddannelsesvejledning	1-7
3	Daily guidance	Daglig vejledning	8-13
4	Work organization	Arbejdstilrettelæggelse	12-17
5	Education	Konference/undervisning	18-20
6	Work climate	Arbejdsklima	21-24
7	Overall Assessment	Øvrige	25-26

Notes: The evaluation scales range from 1 to 6. The individual questions are reported in Appendix Tables G.5 and G.6.

Appendix Table G.3: Questions in Evaluations until 2015, Danish

1	Hvordan vurderer du kvaliteten af introduktionen på uddannelsesstedet?
2	Fulgte du introduktionsprogrammet?
3	Hvordan vurderer du kvaliteten af uddannelsesprogrammet?
4	Svarer indholdet til målbeskrivelsens krav?
5	Svarede uddannelsesforløbet til uddannelsesprogrammet?
6	Har du indfriet checklistens delpunkter?
7	Hvordan var kvaliteten af vejlederens indsats i forhold til din uddannelse?
8	Anvendtes samtaleindholdet (og uddannelsesplanen) i praksis?
9	Hvordan var graden af supervision?
10	Var vejlederen tilstede i tilstrækkeligt omfang?
11	Anviste vejlederen dig uddannelsesrelevante arbejdsområder?
12	Hvordan vurderer du graden af selvstændighed i det kliniske arbejde?
13	Hvordan vurderer du arbejdsbyrden?
14	Var arbejdet tilrettelagt med rimeligt hensyntagen til uddannelsen?
15	Hvordan var vagthyppigheden i forhold til vagtens uddannelsesværdi?
16	Hvordan vurderer du uddannelsesværdien af vagtarbejdet?
17	Hvordan vurderer du uddannelsesværdien af dagarbejdet?
18	Deltog du i forskning/kvalitetsudviklingsarbejde?
19	Deltog du i administrativt arbejde?
20	Deltog du i afdelingens formaliserede undervisning?
21	Underviste du selv?
22	Hvordan vurderer du afdelingens uddannelsesmiljø/prioritering?
23	Hvordan vurderer du uddannelsesstedets samlede uddannelsesindsats?
24	Hvordan vurderer du dit samlede uddannelsesudbytte under ansættelsen?
Text	Vejleder
Text	Uddannelsesansvarlig

Appendix Table G.4: Questions in Evaluations until 2015, English

1	How do you assess the quality of the introduction at the place of education?
2	Did you follow the introductory program?
3	How do you rate the quality of the training program?
4	Does the content correspond to the requirements of the goal description?
5	Did the training course correspond to the training program?
6	Have you met the checklist sub-items?
7	How was the quality of the supervisor's efforts in relation to your education?
8	Was the interview content (and the training plan) used in practice?
9	How was the degree of supervision?
10	Was the supervisor present to a sufficient extent?
11	Did the supervisor instruct you in training-relevant work areas?
12	How do you assess the degree of independence in the clinical work?
13	How do you assess the workload?
14	Was the work organized with reasonable consideration for the education?
15	How was the shift frequency in relation to the shift's educational value?
16	How do you assess the educational value of the shift work?
17	How do you assess the educational value of day work?
18	Did you participate in research/quality development work?
19	Did you participate in administrative work?
20	Did you participate in the department's formalized teaching?
21	Did you teach yourself?
22	How do you assess the department's educational environment/priorities?
23	How do you assess the educational institution's overall educational efforts?
24	How do you assess your overall educational output during employment?
Text	Mentor
Text	Head of Educational Program

Appendix Table G.5: Questions in Evaluations from 2016, Danish

1	Uddannelsesstedet og jeg har afstemt forventninger til uddannelseselementet ved introduktionen.
2	Jeg blev introduceret til de opgaver, jeg skulle varetage.
3	Min hovedvejleder og jeg samarbejdede om at udarbejde min individuelle uddannelsesplan.
4	Mit behov for uddannelsesvejledning er blevet opfyldt.
5	De planlagte kompetencevurderinger er blevet gennemført.
6	Kompetencevurderinger er blevet efterfulgt af feedback.
7	Jeg er blevet tilbudt karrierevejledning svarende til mit behov.
8	Jeg har fået feedback i forhold til min evne til at samarbejde med sundhedsprofessionelle.
9	Jeg har fået feedback i forhold til min evne til at agere professionelt.
10	Jeg har fået feedback i forhold til min evne til at kommunikere.
11	Jeg har fået mulighed for at udvikle mig som leder/administrator og organisator.
12	Jeg har fået supervision svarende til mit behov i det daglige arbejde.
13	De daglige læringsmuligheder er blevet udnyttet.
14	De daglige vejledere har været til at få fat på, når jeg havde behov for det.
15	Arbejdstilrettelæggelsen har tilgodeset, at jeg også har varetaget opgaver, der er relevante for, at jeg har kunnet opnå kompetencerne som angivet i uddannelsesprogrammet.
16	I arbejdstilrettelæggelsen er det blevet prioriteret, at der har været progression i min kompetenceudvikling.
17	I arbejdstilrettelæggelsen er vejledersamtaler blevet prioriteret.
18	Jeg har fået mulighed for at udvikle mig som underviser.
19	Jeg har haft mulighed for at deltage i uddannelsesstedets undervisningstilbud.
20	Jeg har haft udbytte af uddannelsesstedets konferencer.
21	Jeg har oplevet, at der er en gensidigt respektfuld omgangstone på uddannelsesstedet.
22	Jeg har været tryg ved at stille spørgsmål til kollegaer.
23	Jeg har kunnet diskutere svære problemstillinger med mine kollegaer.
24	Jeg har oplevet, at jeg har arbejdet som del af et arbejdsfællesskab.
25	Samlet set har uddannelsesstedets indsats været tilfredsstillende.
26	Mit samlede uddannelsesmæssige udbytte har været tilfredsstillende.
Text	Vejleder
Text	Uddannelsesansvarlig

Appendix Table G.6: Questions in Evaluations from 2016, English

1	The place of education and I have reconciled expectations of the educational element at the time of the introduction.
2	I was introduced to the tasks I had to undertake.
3	My main supervisor and I collaborated on preparing my individual education plan.
4	My need for educational guidance has been met.
5	The planned competency assessments have been carried out.
6	Competence assessments have been followed by feedback.
7	I have been offered career guidance according to my needs.
8	I have received feedback regarding my ability to collaborate with health professionals.
9	I have received feedback in relation to my ability to act professionally.
10	I have received feedback in relation to my ability to communicate.
11	I have had the opportunity to develop as a leader / administrator and organizer.
12	I have received supervision according to my needs in the daily work.
13	The daily learning opportunities have been utilized.
14	The daily tutors have been available when I needed it.
15	The work organization has taken into account that I have also handled tasks that are relevant for me to have been able to achieve the competencies as stated in the training program.
16	In the work organization, it has been prioritized that there has been progression in my competence development.
17	In the work organization, supervisor feedback has been prioritized.
18	I have had the opportunity to develop as a teacher.
19	I have had the opportunity to participate in the educational offer of the educational institution.
20	I have benefited from the conferences of the educational institution.
21	I have experienced that there is a mutually respectful tone of voice at the place of education.
22	I have been comfortable asking questions to colleagues.
23	I have been able to discuss difficult issues with my colleagues.
24	I have experienced that I have worked as part of a working community.
25	Overall, the educational institution's efforts have been satisfactory.
26	My overall educational output has been satisfactory.
Text	Mentor
Text	Head of Educational Program

Appendix G.2: Exit Surveys and Inspector Evaluations

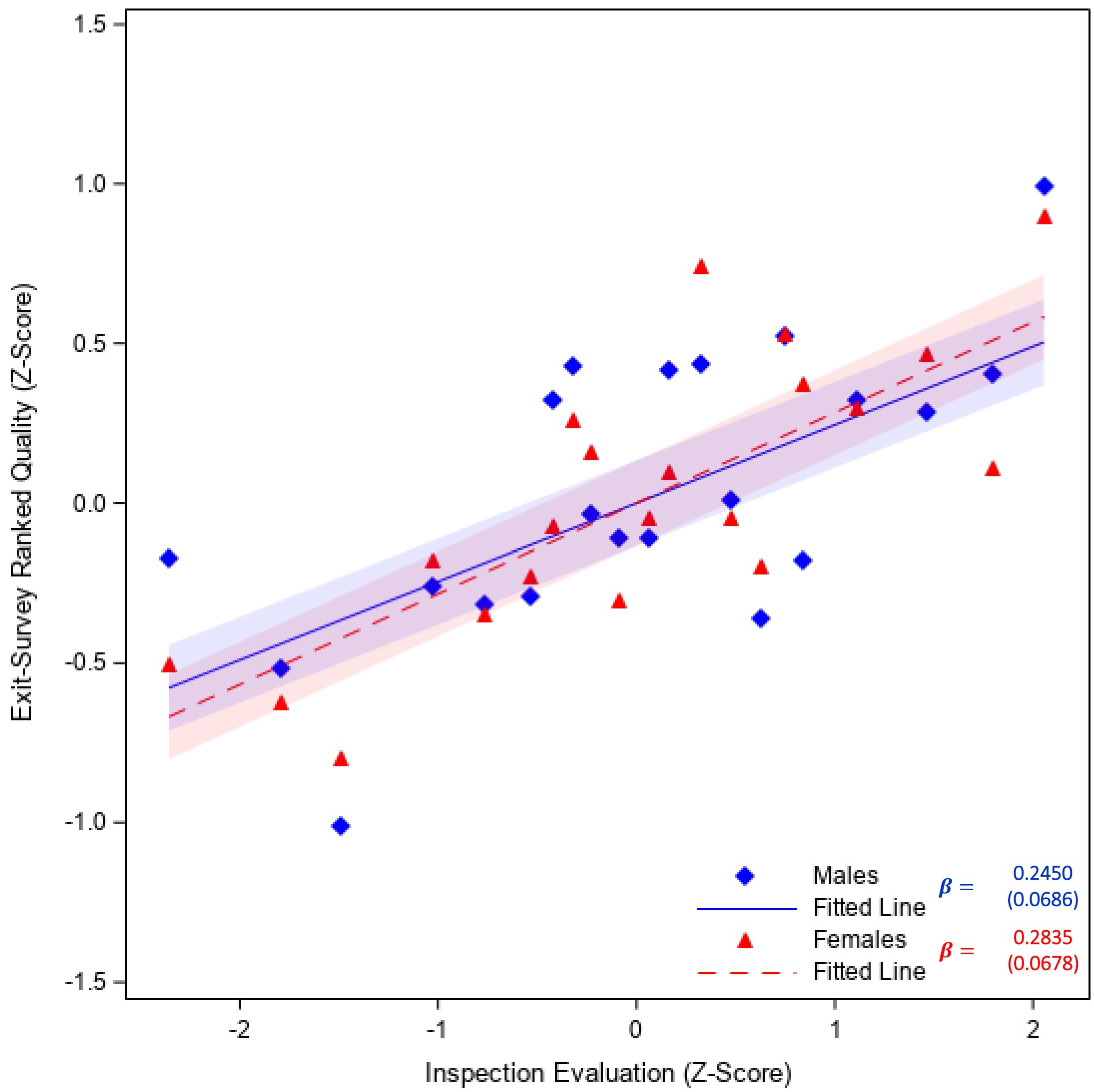
In Section 5.1, we refer to data from external inspections that the National Health Authority (NHA) conducts to assess the quality of the educational programs in hospital departments. In this appendix, we provide details about these assessments and study their correlation with the exit-survey rankings.

The NHA has been conducting external inspections since 1997. Appointed by the NHA, the group of inspectors consists of impartial senior and junior physicians. The inspectors score the hospital department's performance in 16 categories (see panel A of Appendix Table G.7), and each category is scored on a 4-point scale (see panel B of Appendix Table G.7). For our purposes, we use inspectors' overall assessments of a hospital department's internship by summing over all categories. For more details, see: Inspektorordningen Håndbog, Sundhedsstyrelsen, 2016, https://www.sst.dk/da/Udgivelser/2016/Inspektorordningen-Haandbog.

The reports are publicly available on the NHA's website: https://www.sst.dk/da/inspektorrapporter. The NHA servers include inspections from 2013-2022 (where data from 1997-2012 have been erroneously deleted). We hand-code the hospital department IDs for each inspector report in order to link them to our data on the ranked quality from the interns' exit surveys. This provides us with inspector quality assessments of 202 hospital departments (61% of the internship positions).

In Appendix Figure G.1, we study the degree to which inspection assessments are predictive of how interns rank the quality of their internships in the exit surveys. We split the sample into 20 equal-sized bins based on the z-score of the external inspections, where the mean z-score of each bin is displayed on the x-axis. We then plot the average survey-exit ranked quality for each bin (on the y-axis), split by gender.

Appendix Figure G.1: Associations between Exit-Survey and Inspector Evaluations



Notes: This table displays the association between inspection assessments and interns' ranked quality of their internships. We split the sample into 20 equal-sized bins based on the z-score of the external inspections, where the mean z-score of each bin is displayed on the x-axis. We then plot the average survey-exit ranked quality for each bin (on the y-axis), split by gender. We also plot the fitted lines along with 95-percent confidence intervals and report their slopes.

Appendix Table G.7: Structure of Inspector Evaluations

A. Performance Categories for Inspector Assessment

Category	Danish (original)	English (translated)
1	Introduktion til afdelingen	Introduction to the department
2	Uddannelsesprogram	Educational program
3	Uddannelsesplan	Education plan
4	Medicinsk ekspert - Læring i rollen som medicinsk ekspert	Medical expert - Learning the physician’s role as a medical expert
5	Kommunikator - Læring i rollen kommunikator	Communicator - Learning the physician’s role as a communicator
6	Samarbejder - Læring i rollen som samarbejder	Collaborator - Learning the physician’s role as a collaborator
7	Leder/administrator - Læring i rollen som leder/administrator	Leader/administrator - Learning the physician’s role as a leader/administrator
8	Sundhedsfremmer - Læring i rollen som sundhedsfremmer	Health promoter - Learning the physician’s role as a health promoter
9	Akademiker - Læring i rollen som akademiker	Academic - Learning the physician’s role as an academic
10	Professionel - Læring i rollen som professionel	Professional - Learning the physician’s role as a professional
11	Forskning - Uddannelsessøgende lægers deltagelse i forskning	Research - Participation in research
12	Undervisning - som afdelingen giver	Teaching - provided by the department
13	Konferencernes - læringsværdi	The learning value of morning reports
14	Læring og kompetencevurdering	Learning and competence assessment
15	Arbejdstilrettelæggelse - Tilrettelæggelsen tager hensyn til videreuddannelsen af læger	Work organization - The organization takes postgraduate training of doctors into account
16	Læringsmiljøet på afdelingen	The learning environment in the department

B. Assessment Scoring Scale

Score	Danish (original)	English (translated)
1	Særdeles problematisk	Extremely problematic
2	Utilstrækkelig	Inadequate
3	Tilstrækkelig	Adequate
4	Særdeles god	Extremely good