

The (Mis)allocation of Talent Across Medical Specialties: Evidence from Turkey

Veli M. Andirin*

Yale University

September 2026

Abstract

Differences in income across medical specialties do not always compensate for differences in working conditions. A lack of free entry into specialties due to quotas, regulated wages that do not take differences in job amenities into account, or both can lead to almost vertical differentiation across fields: specialties offering both good working conditions and high income (e.g. dermatology) become the most competitive ones to enter, and the very best medical graduates end up in these fields in different contexts (e.g. the US, France, and Turkey). In certain places (e.g. Turkey), the imbalances can be severe: some of the lowest-scoring doctors become brain or heart surgeons, and several specialties are left with a large share of empty residency seats. Using original revealed-preference and survey data from doctors in Turkey, I estimate a structural model of demand for different specialties and characterize an earnings schedule that balances the supply of doctors and talent across specialties at the lowest cost. The resulting earnings schedule differs sharply from the prevailing schedules in different countries and shows that the quotas for highly demanded specialties could be filled even at much lower relative earnings.

*I am grateful to Joseph Altonji, Steven Berry, Barbara Biasi, Leah Boustan, Janet Currie, Philip Haile, Charles Hodgson, Victoria Marone, Mushfiq Mobarak, Cormac O'Dea, and Jesse Shapiro for helpful feedback. I thank Ahmet Ay, the owner of Proders Company, for data and continued support. All errors are my own. E-mail: velimurat.andirin@yale.edu.

1 Introduction

Large variations in compensation and working conditions exist across medical specialties, and the two do not always go in opposite directions: Some specialties offer both better working conditions and income. As a result, they attract the best medical graduates, and there is a sizeable dispersion in the medical ability (e.g. as measured by centralized entrance exams) of doctors across specialties in different contexts. It is not a coincidence that the top medical graduates choose either dermatology or plastic surgery in the US, France, and Turkey, three countries that do not look alike in terms of their healthcare systems.

The reasons that lead to this almost vertical differentiation across specialties might vary across contexts. There is one common reason in all three countries mentioned above, and likely across all healthcare systems where there is a licensing mechanism for medical doctors: The entry into the medical profession, and into its specialties later, is not free. As a result, the supply of doctors to certain highly desirable specialties is usually insufficient to lower their earnings. There can be other sources of imperfection: In certain countries, the government is the largest employer of doctors and has considerable wage-setting power, which has the potential to exacerbate the problem.

In this paper I study the allocation of doctors across medical specialties. I focus on the case of Turkey, where the problems with the allocation look severe: A significant share of residency spots in certain specialties such as emergency medicine, heart and brain surgery remain empty, and these specialties also tend to attract the lowest-scoring students in the nationwide specialty entrance exam. Although the imbalances of demand for different specialties seem large in Turkey, I show that certain features of the allocation of doctors across specialties in this context are also present in other countries. I investigate the determinants of doctors' specialty choice in Turkey using unique data on doctors' preferences and expectations, and characterize the earnings schedule that can achieve a more balanced distribution of doctors across specialties. I show that while this schedule balances the overall demand across different specialties, it might alter the gender composition across fields in unintended ways. I then discuss how this counterfactual earnings schedule compares to the private and government sector earnings in Turkey, and to the overall earnings in the US.

I start by laying out the details of the institutional context. The first step in Turkish medical training is the medical school, which lasts six years. Doctors who would like to specialize in a field take a centralized exam after they finish the medical school, and their assignment to residency programs is based on their scores in this national exam. Residency programs last around four years, with some specialties, such as surgical ones, requiring longer training as in other contexts. The government organizes the residency match, and also leads the process of determining the number of seats available in each program. In addition to its role in planning the match, the government is

also the largest employer of doctors in Turkey, and does almost the entire healthcare spending in the country.

I then describe the data in greater detail. The main data in this paper come from a match simulation platform: Doctors in Turkey enroll in a match simulation platform before the official deadline for submitting their rank-order lists. They submit their tentative lists to this simulation platform. The platform simulates the match taking these tentative lists as its inputs, and predicts the residency program assignments. This process happens every few minutes, and as I show, the final predictions on this platform closely track the real match outcomes. I obtain snapshots of doctors' rank-order lists and their predicted assignments from this platform. I supplement it with an original survey of doctors' expectations about income and working conditions in different specialties in Turkey, as well as official data on match outcomes in Turkey, France, and the US to enable international comparisons.

Using these data, I document facts about the allocation of doctors across specialties in Turkey, its relation to perceived earnings and working conditions in the government and the private sector, and how these compare to other countries. I show that the link between expected weekly hours and compensation is weaker in the government compared to the private sector. I document that weekly hours are tightly correlated with perceived difficulty of a specialty, but the relationship is imperfect, highlighting the importance of other factors. I present descriptive evidence on the gender gap in preferences for different specialties. I show the similarity of the Turkish residency entrance exam and the licensing exam in the US (USMLE) using the performance of Turkish doctors in answering USMLE questions that were translated as a part of this study. I finally document the rankings of medical specialties using exam scores in Turkey, France, and the US, and discuss the similarities and differences in these three contexts.

After these descriptive findings, I build and estimate a model of doctors' specialty choices. The fact that the doctors observe what was available to them based on their exam scores at the time of choice allows me to define their choice sets clearly, which often poses a challenge in estimating preferences for hospitals (or schools) as it is difficult to disentangle preferences and choice constraints from submitted lists or aggregate match outcomes. Another common challenge in the literature on estimating preferences for medical specialties is the correlation between unobserved amenities at the specialty level and earnings. The survey data that I collect on weekly hours and perceived difficulty allow me to control for these factors when estimating the preferences.

I estimate the preferences separately by gender, using a two-step procedure following Berry et al. (2004). The first step produces estimates of mean utilities for each specialty, which shows clear demand differences for specialties across genders. These estimated mean utilities are then used in a second step to learn about the impact of expected earnings, working hours, and perceived difficulty in determining the choice of specialties. The estimates show that both genders derive a

disutility from longer hours and increased difficulty, but women have a stronger distaste for longer hours.

I use the estimates from the model to solve the problem of a planner that aims to fill all residency positions in each specialty while minimizing the total payments. This objective is motivated by the fact that the seats are determined by a process that takes the needs of the teaching hospitals and long-term goals of the government into account. I show that the seats in the most desirable specialties would still be filled even if their earnings were reduced significantly. Moreover, the resulting allocation of doctors into specialties also leads to very similar mean exam scores across fields. I also emphasize that aiming to fill residency positions via changes in earnings can lead to other imbalances: For instance, while dermatology seats can still be filled despite much lower incomes, it would become an almost women-only profession because of women's strong preference for the non-pecuniary attributes of this specialty. I finally show that the resulting earnings schedule is only weakly correlated with either of the private sector earnings in Turkey or with the earnings across specialties in the US, suggesting that the root cause of the current imbalances in the distribution of doctors across specialties may not be the heavy government presence in the Turkish healthcare sector.

Based on the results in this paper, it might be tempting to suggest additional roles to governments such as setting earnings across specialties more actively, especially given the lack of free entry into medical specialties and the interest from policymakers in shaping the medical workforce. For instance, a recent Congressional Report (Congressional Research Service 2025) in the United States suggests that the government could assume a more active role in determining the composition of specialists beyond being mainly a source of funding for graduate medical education. However, the very example in this paper comes from a context where the government has significant levers in shaping the medical workforce, and yet, the resulting allocation is perhaps more counterintuitive than the one in cases where the private sector plays a bigger role.

This paper connects to the literature on the specialty choices of medical doctors. A closely related paper that studies the specialty choice of physicians in the context of the US is Nicholson (2002). Nicholson (2002) constructs a measure of expected earnings by relying on estimated probabilities of being matched without observing the choice sets, and finds that income plays a key role in specialty choices. Another closely related paper that is more recent is Gottlieb et al. (2025). Gottlieb et al. (2025) study physician earnings, and how it impacts the allocation of talent across specialties in the US, using an estimation strategy based on aggregate shares following Berry (1994). The strategy used in this paper for estimating doctors' preferences for specialties differs from these previous studies in several ways: The assignment mechanism, and the data on doctors' exam scores as well as the rank-order lists allow me to observe what was available to each doctor at the time of their choice. I also elicit doctors' expectations about working hours and other

factors influencing the difficulty of a specialty separately, which allow me to control for important but typically unobserved factors that might be correlated with the expected earnings. Individual-level data on exam scores and the preference estimates then allow me to simulate the impact of counterfactual income schedules.¹

My work also relates to the literature on gender sorting across occupations. A closely related paper is Wasserman (2023), who studies how working conditions affect gender sorting across specialties. I add to her results by providing structural estimates of the male and female preferences for different specialties, and how they tradeoff income, working hours, and other working conditions. My results also enable me to demonstrate how alternative income policies would impact gender sorting across medical specialties.

Also related to this paper is the work by Buehler et al. (2026), who provide a comparison of physician incomes across countries. In this paper, I provide evidence on talent allocation across specialties in different countries. I also document how the expected earnings and working hours in the context of Turkey correlate with actual earnings and working hours in the US.

The rest of this paper proceeds as follows. In Section 2 I provide a more detailed overview of the institutional setting and the data sources. Section 3 provides descriptive facts about the demand for different specialties, talent allocation, expected earnings and working conditions in the context of Turkey, and also presents international comparisons. In Section 4 I discuss the model of specialty choice, identification, estimation, and the results. In Section 5 I discuss a counterfactual income schedule and its implications. Section 6 concludes.

2 Institutional Setting and Data

2.1 Institutional Setting

Medical school in Turkey lasts six years. Upon completing the medical school, doctors need to work in underserved areas of the country in a program called the compulsory medical service.² After this step, doctors who would like to specialize in a field take the centralized residency placement exam.

The exam and the placement process take place twice a year, in the spring and in the fall. There

¹Other related papers studying doctors' specialty choice include Sivey et al. (2012), who use a discrete choice experiment and quantify the importance of pecuniary and non-pecuniary attributes. In a follow-up paper, Méndez et al. (2025) show that data in discrete-choice experiments are correlated with future choices. Other related studies that use experiments include Li (2018) and Attema et al. (2023), who show that physicians' altruism is related to expected specialty choices.

²See Andirin 2025 for a more detailed discussion of the compulsory medical service in Turkey. The official duration of the service varies depending on where the doctor serves, however, doctors can influence their length of service by taking the residency exam and enrolling in a residency program. As noted in Andirin 2025, most doctors express that they intend to serve around a year during this service.

are 41 different specialties that the doctors can consider when applying for residency programs. The programs typically last around four years. Appendix Table 1 provides a list of programs along with the required years of training. Candidates are placed into programs via a centralized assignment mechanism that uses multi-category serial dictatorship. (See Balinski and Sönmez 1999 for a discussion of this mechanism). The number of available seats in each program is determined by the Ministry of Health, taking into account both the demands from individual institutions and the total supply of doctors to different specialties (the Official Gazette 2022). In each placement cycle, there is a supplementary placement round where the applicants who were not matched at the end of the first round and the remaining seats are matched, using the same mechanism.

In addition to its role in determining the supply of residency spots, the government is also the largest provider of healthcare in Turkey. In 2023, approximately 78% of the healthcare spending was by the government, followed by the out-of-pocket healthcare spending by households, which comprised roughly 18% of the spending (Turkish Statistical Institute 2023). The lack of a significant role of private health insurance is striking: Only 2.5% of the total health spending was by private health insurance (Turkish Statistical Institute 2023). The largest employer of healthcare workers is the government, too: 80% of all doctors in Turkey are government employees, and the number is similar for specialists, with 68% of them employed by the government.

The compensation for specialists in the government sector consists of a base compensation, which does not depend on the specialty, and bonus payments, which can vary between specialties, but also between hospitals. The base pay can also differ based on geographic location, and specialists in less developed areas receive a higher base pay, but the differences are small.³

To the best of my knowledge, there is no publicly available source that includes the final compensation for doctors from different specialties in Turkey in the government sector. A similar lack of data availability is also present for the private sector. There are no official statistics on the earnings of the doctors from different specialties in the private sector. Job posting websites contain very limited information as the job search process for doctors often occurs through other channels. Another challenge in obtaining data on earnings for doctors in the private sector is that these data might be missing even for the government. Recent news headlines indicated that the Ministry of Treasury in Turkey is preparing for imposing a flat fee for certain occupations with private offices, such as doctors and lawyers, due to consistent underreporting of their earnings (NTV 2025). To overcome these challenges, I included questions regarding the expected earnings in the private and the government sector in my surveys.

³A starting specialist in the capital, Ankara, can expect to receive a base compensation around \$2800 per month, and a starting specialist in one of the least developed districts in the country in the city of Hakkari can expect to receive a base compensation around \$3000. Similar to other government employees, there can be small changes in the base pay also depending on the number of children, the working status of the spouse, and so on. Similar to other government employees, the salary increases based on the years served.

2.2 Data

My main data come from a private company that simulates the residency match in Turkey. In the weeks leading to the match, doctors who take the residency exam and plan to enter the match enroll on a match simulation portal. They submit their rank-order lists to this platform, and based on their verified exam scores and the submitted rank-order lists, the platform runs the match, and shows the doctors where they are predicted to be assigned. This process happens every few seconds.

The data used in this paper come from the main round of the match that took place in Fall 2025, where 29766 doctors took the exam to compete for 10429 available seats. At the end of the main round, 8435 doctors were placed in a residency program. The supplemental round that took place later increased the number of matched doctors by 43.

The number of doctors who used the platform in this round of the match was 10,124. This is smaller than the number of doctors who took the exam, but larger than the number of doctors who were placed. This is expected since there would not be much point to using the platform if a doctor does not consider being placed as a meaningful possibility.

It is challenging to establish what share of doctors who enroll in a residency program use the platform, since the eventual placement outcomes of the doctors are not verifiable. However, it is informative to consider how platform-predicted outcomes track the real outcomes at the residency program level. There were 2,946 residency programs in total. In Appendix Figure 1, I show the distribution of the differences between the platform-predicted and realized cutoffs for entry into different residency programs, measured in units of the specialty exam score. For the majority of the programs, the difference is less than one point in absolute value, which is a tenth of the standard deviation of exam scores. This suggests a wide, and mostly truthful use of the platform. At the end of the simulation process on the platform, 7964 doctors were predicted to be placed. Based on the closeness of the real and predicted outcomes at the program level, this suggests that as much as 94.4% of the doctors who were matched could have been using the platform.

I obtained four main types of data from this platform. The first type of data are the inputs and outputs of the match simulation: I obtain snapshots of doctors' rank-order lists on the platform and their predicted assignments every five minutes.

The second type of data are original surveys about doctors' expectations for earnings, working hours, working conditions, and private sector employment chances in different specialties. I asked the users of the platform their expected monthly earnings both in the government and in the private sector if they were placed in their top-choice specialty on the platform.⁴ I then asked their expectations about weekly working hours in both sectors. In addition to working hours, I asked

⁴A limitation is that I did not ask their expectations about income in other specialties in these pop-up questions. This is due to the fact that both the number and the format of questions that I could ask through a platform that the users pay for were limited.

doctors to think about other factors (such as the risk of malpractice liability, night floats, high-stakes procedures) to assign each specialty a difficulty score on a scale from 1 to 9, with 9 being the highest level of difficulty. Through these pop-up questions, I also asked doctors their perceived chances of working in the private sector.

I also obtained data on the characteristics of the doctors and the programs. Data on doctor characteristics include their gender, exam score, and the medical school they graduated from. For the residency programs, I obtained their location and type (e.g. whether it is university-based or hosted by a city hospital).

Finally, I assessed the performance of Turkish doctors on the licensing exam in the United States (USMLE), since the performance of doctors on the latter has been associated with improved patient outcomes (Norcini et al. 2024). To do so, I translated a subsample of USMLE questions from various topics to Turkish, and each doctor was given a set of three randomly chosen questions from this subsample.

I supplemented these data with official placement statistics following the matches, obtained from the Student Selection and Placement Center (SSPC), for both the main and the supplemental round of the match. These include the number of available seats in each program, the number of filled seats, and the realized cutoffs.

For cross-country comparisons, I obtained data on exam scores across specialties in the US from the National Resident Matching Program (National Resident Matching Program 2024) and in France from the official ECNi assignment results published in the Journal Officiel (Centre National de Gestion 2023). I also obtain data on earnings and working hours across specialties in the US from Gottlieb et al. (2025).

3 Descriptives

3.1 Capacity, Empty Seats, and Talent Distribution Across Specialties

I present the number of available seats and the share of empty seats across specialties in Figure 1. The figure makes it clear that several specialties face severe problems in terms of filling their available seats, and these are usually the surgical specialties. This large share of unfilled seats is hardly explained by an abundance of available spots, the majority of surgical specialties with a large share of empty seats offer a relatively small number of spots, suggesting that the underlying problem is due to a lack of demand.

This lack of demand for certain specialties is also illustrated in the distribution of exam scores shown in Figure 2. Specialties that have the largest share of unfilled seats also tend to attract the lowest-scoring students in the exam. For instance, in addition to the problem of having nearly thirty percent of its seats unfilled, cardiac surgery also attracts some of the lowest-scoring students

in the residency exam.

3.2 Income, working hours, and perceived difficulty

Figure 3 shows the relationships between income, working hours, and perceived difficulty across specialties. The number of respondents to survey questions about specialty characteristics was 1,043. As a result, since there are 41 specialties, the number of responses per specialty is limited especially for small fields. To make the plots more informative, I aggregate specialties as described in Appendix Table 1, and refer to these groupings as specialties in Figure 3.

As can be seen in Panels A and B, there is a positive relationship between the expected monthly income and working hours in both the government and the private sector, but the relationship is much steeper in the private sector. The overall level of expected earnings in the private sector is higher. Some specialties are outliers in both sectors, for instance, the expected earnings in Plastic Surgery are relatively high compared to specialties with similar working hours in both sectors, which is suggestive about the reason for its place at the top of the exam score distribution in Figure 2.

Panel C shows the relationship between the average weekly working hours and the difficulty levels of specialties. To construct average working hours, I take a weighted average of the weekly hours in the government and private sector, where the weight on the private sector hours is given by the average expected probability of being employed in the private sector as reported by the survey respondents. Although the difficulty question explicitly nudged the respondents to think about factors other than the weekly working hours when scoring the difficulty levels of specialties as explained in Section 2.2, there is nevertheless a very tight relationship. However, certain specialties appear to be relatively easy given their working hours. For instance, Urology has a meaningfully smaller difficulty score than Orthopedics despite having similar weekly working hours, perhaps reflecting lower-stakes procedures on average.

In addition to asking doctors to score specialties in terms of their difficulty, I also asked open-ended questions where doctors commented on the positive and negative aspects of the specialties they chose. For each response, I built flags indicating whether the response contains a keyword about a certain job characteristic, such as the working hours, overall stress levels, income, and professional satisfaction. The number of responses to these open-ended questions was lower (131 respondents in total), and as a result, there are only a handful observations per specialty even when using the coarser categorization in Appendix Table 1. For this reason, Figure 4 aggregates the responses at the level of surgical and non-surgical specialties.

Working hours is the leading theme when it comes to the negative aspects of surgical special-

ties, and it rarely gets mentioned in responses to the question about positive aspects.⁵ On the other hand, for non-surgical specialties, it is the most frequently stated positive side (as doctors mention that they find the working hours lighter for these specialties). The risk of being liable for malpractice is the second most frequently mentioned negative aspect of surgical specialties. In contrast, the mention of malpractice appears more frequently among responses to positive aspects for non-surgical specialties (as the respondents usually mention the low risk of liability in these fields). Another distinguishing aspect between the surgical and non-surgical specialties is the mentions of stress, which frequently appear as a negative aspect of surgical specialties, and rarely gets mentioned as a positive aspect. For non-surgical specialties, it frequently appears in responses about positive aspects, indicating lower levels of stress in these fields. Professional satisfaction is also a distinguishing factor between the surgical and non-surgical specialties as it is the most frequently mentioned positive aspect for surgical specialties.

In Appendix Table 2, I show the relationships between the themes that appear frequently in open-ended questions and the difficulty scores of specialties. To construct this table, I regress an indicator for whether a response (about positive or negative aspects of the chosen specialty) contains a keyword from a given theme (e.g. working hours) on the average difficulty level of the specialty group that the respondent's specialty belongs to. Considering the fact that surgical specialties are often scored as more difficult, the table echoes the findings from Figure 4 that compared surgical and non-surgical specialties: As the difficulty of a specialty goes up, responses about negative aspects are more likely to include keywords related to stress, malpractice, and working hours, and responses about positive aspects are more likely to include professional satisfaction.

3.3 Test scores and specialty entry by gender

It is also important to consider whether preferences for specialties vary by gender, as potential policies that aim to alter the distribution of talent across specialties might also have an impact on the gender composition across fields.

The structural model that will be presented in Section 4 studies the preferences separately by gender. In this section, I present descriptive evidence suggesting different preferences for various fields by men and women, by looking at the shares of doctors entering each specialty separately by gender. One concern with this approach is that the entry into a residency program is constrained by capacity and exam scores. However, Panel A of Figure 5 shows that the distribution of exam scores looks similar for both genders, suggesting that looking at shares can be informative.

Panel B of Figure 5 shows the share of candidates entering each specialty, separately by gender. The share of male doctors entering surgical specialties such as heart and brain surgery is relatively

⁵For certain surgical specialties such as ophthalmology, working hours were mentioned as a positive aspect.

higher compared to females. However, there are surgical specialties where the share of females entering the specialty is higher than the corresponding share among males, such as OB-GYN.

It is worth noting that there are other shortcomings of comparisons based on the shares, even if the exam score distribution is similar across genders. For instance, men and women might value differently the characteristics of programs such as their type, location and distance to where they finished the medical school. The structural model in Section 4.3 will provide estimates of the mean utilities for each specialty by gender, netting out the effect of such characteristics, and also taking into account the exam scores and capacity constraints.

3.4 Benchmarking the Turkish specialty exam

The results presented so far indicate a large dispersion of average exam scores across specialties in Turkey, with most surgical specialties clustered at the bottom of the distribution. Before studying the impact of alternative income schedules on exam score distribution across specialties, it is useful to ensure that the residency exam in Turkey captures some meaningful signal about medical knowledge. I do so by establishing the similarity between the exam in Turkey (TUS) and the USMLE. An advantage of the latter is that previous studies demonstrated the link between higher USMLE scores and better patient outcomes (Norcini et al. 2024).

For this purpose, I chose sample questions from individual USMLE steps in a way that ensures coverage across core topics in the USMLE guidelines. I then asked doctors to answer three randomly chosen questions from this pool of sample questions. I then estimated:

$$Correct_{iq} = \alpha_q + \beta TUSScore_i + \varepsilon_{iq}$$

where the outcome variable is whether doctor i answered the question q correctly, α_q is a question fixed effect, and $TUSScore_i$ captures the doctor's score in the Turkish exam.

The results are presented in Table 1. There is a statistically significant, positive relationship between the Turkish exam scores and performance on USMLE questions.

Converting a standard deviation increase in the Turkish exam scores into a standard deviation increase in the number of correctly answered USMLE questions is challenging, because there is no publicly available data on how correct answers in the USMLE are converted into USMLE scores. However, a psychometrics paper written by NBME authors (Raymond et al. 2019) studies a high-stakes physician licensing test with around 320 questions, which is very likely to be USMLE Step 2.⁶ According to that paper, the standard deviation of the percentage of questions answered correctly is around 8%, implying that the standard deviation of the number of questions answered correctly is around 24. Assuming that an increase in TUS performance leads to a similar increase

⁶This insight is from a blog post by a medical doctor (Carmody 2020).

in the number of questions answered correctly across USMLE steps, approximately 82% of a standard deviation in correctly answered USMLE questions can be explained by a standard deviation increase in TUS scores.⁷

3.5 Talent distribution across specialties in other countries

It is also instructive to compare the distribution of talent across specialties in different countries. In this section, I show how the relative rankings of specialties in terms of the exam scores of incoming residents vary across France, Turkey, and the US. This task is not straightforward, as the content of exams used for specialty entrance, and the definitions of specialties can vary across contexts.

Section 3.4 has shown that the exams in seemingly very different contexts can be similar. To deal with the second challenge, I map the specialty programs in Turkey and France to those in the US, as described in Appendix Table 3. Although the mapping is clear for certain specialties (such as Dermatology), it is less clear for others. The specialty categories in the US are coarser, for instance, Cardiology and Pulmonology are subspecialties of Internal Medicine and they require fellowship training after the completion of an Internal Medicine residency. I map such specialties to their most direct counterparts in the French and Turkish systems for simplicity.

For the US, I use average Step 2 scores across specialties in 2024 taken from NRMP (National Resident Matching Program 2024).⁸ I assign higher ranks to specialties with higher average scores. For Turkey, specialties are ranked based on the average TUS score. For France, I rank specialties based on the average rank of the incoming resident in the ECN test in 2023.⁹

Figure 6 shows the rankings of specialties based on exam scores in these three countries. There are similarities at the top. Dermatology and Plastic Surgery are among the top-ranked specialties in all three countries. There are similarities at the bottom of the distribution across the three countries as well: Pediatrics, Emergency Medicine, and Family Medicine all appear near the bottom of the distribution in all cases. This similarity at the bottom is particularly striking, given that these centralized exams typically measure a more general knowledge of medicine rather than specialized knowledge, and the entrants into specialties where a more general knowledge can be useful seem to have less of it.

A key difference across countries is the relative position of surgical specialties. Major surgical fields such as General Surgery, Vascular Surgery, and Neurosurgery appear either in the middle or in the lower half of the exam rank distribution in France and Turkey, but they are usually in the upper-middle spots in the US.

⁷The standard deviation of TUS scores is 10 by construction of the scoring formula (OSYM 2025).

⁸Since 2022, Step 1 scores are no longer published, and the exam has become pass/fail only.

⁹2023 was the last year in which there was a common test score across all specialties. With the introduction of a new system in 2024, there are now specialty-specific exam scores.

There are important differences as well: Orthopaedic Surgery shares the top rank in the US with Dermatology, however, it is among the middle-ranked specialties in France and Turkey. Overall, surgical specialties are ranked higher in the US, and fields such as Neurosurgery and Orthopaedic Surgery are in the bottom half in Turkey and France.

What explains the similarities and differences across countries? Figure 7 compares the income and working hours across specialties in Turkey and in the US. To construct this figure, for each NRMP specialty, I find the corresponding Turkish specialty group as described in Section 3.2. I present the mapping between NRMP specialties and Turkish specialty groups in Appendix Table 4. When multiple NRMP specialties map to the same Turkish group, I take the unweighted average of earnings or working hours across these specialties.

In Panel A, I compare the relative earnings of each specialty group across the two countries, where earnings in each country are expressed relative to the cross-specialty mean. There is a meaningful positive correlation: specialty groups that earn more relative to the mean in the US also tend to earn more in Turkey. However, the relative income dispersion across specialties is considerably wider in the US. In Panel B, I compare the weekly working hours and find an even tighter correlation. This is also valuable for demonstrating that the survey data on working hours captures meaningful variation across specialties.

Differences and similarities in terms of the exam rankings of specialties across countries can hardly be explained by a single factor, as can be seen in Figures 6 and 7. For instance, Dermatology has relatively high earnings compared to the cross-specialty mean in the US, but this is not true in Turkey. Yet both countries share the same pattern of Dermatology attracting the highest exam scores. Likewise, surgical specialties such as Neurosurgery, Cardiac Surgery, and Obstetrics and Gynecology have similar earnings relative to the cross-specialty mean in both countries, but they attract much higher-scoring applicants in the US. These patterns highlight the need for a structural model of specialty choice that can account for multiple factors simultaneously.

4 An Empirical Model of Residency Program Choice

In this section I consider a model of residency program choice by doctors. I use the estimates from the model to quantify preferences for different specialties by gender, to characterize how doctors trade off pecuniary and non-pecuniary characteristics, and finally to solve for the compensation scheme that fills all residency seats at the lowest cost.

4.1 Model

I model the utility that a doctor i gets from a residency program j as:

$$U_{ij} = \delta_{s(j)}^g + \beta^g X_{ij} + \varepsilon_{ij}$$

where $\delta_{s(j)}$ refers to the component of utility that comes from receiving training in the specialty $s(j)$ offered by the program j . X_{ij} is a vector that collects the distance between the city where the doctor graduated from the medical school and the program's city, the development level of the program's city, and dummies for the type of the hospital where the program is offered. ε_{ij} is a Type-I extreme value shock i.i.d. across doctors and programs. Importantly, I allow the coefficients to vary by the gender of the doctor, indexed by g . I normalize the utility from the outside option of not being matched to any program to zero.

I then decompose the specialty-specific component of the utility, $\delta_{s(j)}$, into a part that is common across specialties, parts that are explained by the monthly expected income, weekly working hours and the difficulty level of the specialty, and finally an unobserved taste for the specialty:

$$\delta_s^g = \gamma_0^g + \gamma_1^g \text{income}_s + \gamma_2^g \text{hours}_s + \gamma_3^g \text{difficulty}_s + \eta_s^g$$

An important feature of the data I have is that the doctors know, based on the rank-order lists submitted by all doctors, the options that are feasible to them. The platform shows the cutoff scores for all programs. As a result, I assume that the feasible program ranked the highest by the doctor, μ_{ij}^* , maximizes her utility among the set of all feasible programs, $\mathcal{S}_i$ given the doctor's score ¹⁰:

$$\mu_i^* = \arg \max_{j \in \mathcal{S}_i} U_{ij}$$

4.2 Estimation and Identification

Under the distributional assumption on ε_{ij} , we have:

$$Pr(j = \mu_i^*) = \frac{\exp(\delta_{s(j)}^g + \beta^g X_{ij})}{1 + \sum_{j' \in \mathcal{S}_i} \exp(\delta_{s(j')}^g + \beta^g X_{ij})} \quad (1)$$

Using equation 1, I estimate the parameters δ_s and β^g via maximum likelihood. The estimated δ_s parameters can then be used in a second stage to recover the γ parameters by estimating:

$$\delta_s^g = \gamma_0^g + \gamma_1^g \text{income}_s + \gamma_2^g \text{hours}_s + \gamma_3^g \text{difficulty}_s + \eta_s^g \quad (2)$$

Identification of the coefficients on observables requires orthogonality with the unobserved taste term. In particular, since the counterfactual exercise will aim to construct an alternative earnings schedule, it is crucial to estimate γ_1 consistently. In most settings, the correlation between

¹⁰A similar model of choice appears in Bobba et al. (2021), who studies teacher sorting across schools in Perú.

income and the unobserved taste is concerning, since for instance specialties that offer higher wages can be unobservably less desirable (e.g. due to harder working conditions), leading to a negative association between the two. I argue that this is less concerning here, as the data on weekly working hours and specialty-level difficulty assessments aim to capture precisely such typically unobserved factors.

The data for estimating equation 2 comes from the surveys that were described in Section 3.2. As discussed there, the availability of data varies across specialties, with certain specialties having a limited number of observations. When constructing expected income, expected weekly hours, and overall difficulty levels for equation 2, I follow the approach for constructing empirical Bayes posterior means outlined in Walters (2024). Specifically, I group specialties as surgical and non-surgical, and treat the mean parameters for a given specialty as drawn from a normal distribution that is common to all specialties in the same group. This approach has the advantage of using both the survey data for a given specialty, but also data from other specialties in the same group. The weight on the latter goes up when the specialty’s own data is relatively noisier.

Using this approach, for each specialty, I construct the posterior means of income and weekly hours for the private and the government sector as well as the difficulty level of the specialty and the probability of working in the private sector. Finally, to construct the expected income and weekly working hours in a given specialty, I take weighted averages of posterior mean income and working hours in the government and private sectors for each specialty, where the weight on the private sector is given by the posterior mean probability of private sector employment for that specialty.

4.3 Results

Table 2 shows the coefficient estimates from equation 1. Figure 8 visualizes the relative utility of different specialties by gender, where for each gender, I normalize the lowest mean utility for specialties to zero. An advantage of comparing the mean utilities rather than looking at the shares of incoming residents for each gender across specialties is that the former takes into account the availability of seats (both due to quotas and also due to exam scores) and nets out the effect of distance, hospital type, and the location’s quality.

There are several takeaways from these estimated mean utilities. First, the preferences are similar overall between the two genders, and specialties with better working conditions and/or income as described in Section 3.2 in general appear on top, and specialties with harder working conditions and/or lower income appear at the bottom. However, there are some visible differences between the preferences of the two genders as well. The preference of females for psychiatry (both adult and children/adolescent specialties) and foundational science specialties (such as microbiology, biochemistry) appears stronger than males’ preferences. Moreover, fields such as cardiology,

cardiovascular surgery, and brain surgery are at the bottom of the preference ranking for females. Males do not have a high mean utility for these fields either, but their ranking is relatively better (e.g. cardiology is in the middle of the distribution, and cardiac and brain surgery appear slightly above the very bottom).

I then present the results from estimating equation 2 in Table 3. All coefficients have the expected signs, with both genders deriving increased utility from higher income and disutility from higher working hours and difficulty. The relative willingness to pay of females for lower weekly hours is higher than that of males, consistent with the results in Wasserman (2023). The relative disutility for other difficult aspects of specialties is similar across genders.

5 Counterfactual

5.1 Solving for the Earnings Schedule That Fills the Empty Seats

As discussed in Section 3.1, certain medical specialties in Turkey struggle in terms of attracting medical graduates, and as a result, a large share of their seats are left empty. This is also reflected in the average exam score distribution across specialties: Those specialties with the highest share of empty seats are also the ones that have the lowest average specialty entrance exam scores.

Since the number of available seats is determined by a process where long-term planning goals of the Ministry of Health and the demand from individual institutions is taken into account, as mentioned in Section 2.1, in the counterfactual analysis in this section, I consider the problem of a planner that is trying to fill these seats. In other words, I take the existing quotas as reflecting genuine needs for medical residents, and then ask how the planner can recruit the desired number of residents in each specialty at the lowest cost possible. I also take the current matching mechanism, multi-category serial dictatorship, as given.

Formally, I solve:

$$\min_{\{w_s\}_s} \sum_s w_s n_s(\{w_s\}_s)$$

subject to:

$$n_s(w_s) = Q_s \quad \forall s$$

where w_s is the expected income in specialty s , $n_s(w_s)$ is the number of specialists in specialty s as a function of the income of all specialties (omitting from the notation the dependence on other characteristics and shocks for simplicity), and Q_s is the total quota of specialty s .

I solve this problem by an algorithm where specialties "bid" for higher fill rates. More precisely, I first set the income of all specialties to a very low amount (10% of the current lowest

income across all specialties), run the match once candidates have re-ranked all programs under this new income schedule (using a fixed draw of unobserved taste parameters ε_{ij} and using the residuals $\hat{\eta}_s$ from equation 2), and record the resulting assignment rates. Then, in each iteration, I increase the income for each specialty that has not yet reached its quota incrementally. This procedure is in line with the bidding procedure in Kelso and Crawford (1982) and the generalized deferred acceptance algorithm in Hatfield and Milgrom (2005); and was recently applied in Bobba et al. (2021).

5.2 Resulting Income Schedule, Talent Allocation, and Gender Composition

Panel A of Figure 9 shows the relationship between the resulting counterfactual income schedule and the current income at the level of specialty group for ease of comparison. As expected, to achieve high fill rates, the earnings in most surgical specialties have to go up considerably. However, this does not apply to surgical specialties that already experience high fill rates in the status quo. In fact, surgical fields such as plastic surgery and ophthalmology could fill all of their seats even if they offered significantly smaller earnings, which can be explained by the high desirability of these specialties in other dimensions. Perhaps the most striking example of a specialty that would still fill every seat despite a reduction in income is dermatology. The results show that even with a decline of 32% in earnings, all dermatology seats could be filled. On the other hand, the specialties that would need an earnings increase are not limited to certain surgical ones: Fields such as internal medicine or acute care would need to experience a large increase in earnings to fill most of their seats.

Panel B of Figure 9 shows the share of female doctors in each specialty group in the status quo and in the counterfactual scenario. The range between the minimum and maximum share becomes higher under the counterfactual income schedule. In particular, specialties that were widely preferred among women in the status quo such as dermatology and psychiatry become much more female-dominated under the counterfactual. This suggests that an income schedule that targets high fill rates as the objective might lead to unintended consequences along other dimensions. Differences in preferences across the two genders for different aspects of specialties are at play here: Since the female preferences for non-pecuniary aspects of certain specialties are strong, even if these specialties experience a large income decline, women might be willing to stick with them. On the other hand, since the distaste of men towards surgical specialties is weaker, they leave non-surgical specialties in greater proportion when surgical fields experience an earnings increase.

Table 4 presents the full results at the level of specialty, including the fill rates and average exam scores in addition to earnings and gender composition discussed earlier. It is worth noting that, although average exam score levels in individual specialties were not targeted directly in the

counterfactual problem, aiming for high fill rates in all specialties leads to a balanced distribution of exam scores across fields. This is intuitive, since both the empty seats and low average exam scores for certain specialties reflect low demand, and increasing the compensation in specialties with a large share of empty seats also increases the average exam scores in these fields.

As discussed in Gottlieb et al. (2025), it is crucial to think about the extensive margin responses to different income allocations across specialties. A cut across the board might induce potential physicians to prefer other occupations which would decrease the overall supply of doctors. In the context I study, the goal of filling every seat in every specialty means that the average income of a specialist, weighted by the number of doctors enrolling in each specialty, has to go up from \$5,172 to \$6,484, despite large declines in the earnings of certain specialties, which makes negative impacts on the overall supply of doctors less likely.

5.3 Comparing the Counterfactual Earnings to the Government, Private, and US Earnings

One might wonder whether the actual earnings would look more similar to the counterfactual earnings outlined in the preceding section if the private sector played a bigger role in the healthcare industry. Figure 10 shows how the counterfactual earnings schedule relates to the earnings in the government and private sector in Turkey, and also to the earnings across different specialties in the US. To facilitate the comparisons, I use the specialty groups described earlier in Section 3.2 and in Section 3.5.

Panel A shows the correlation between the relative expected earnings across specialties in the government sector and the relative earnings across specialties under the counterfactual earnings schedule. Panel B repeats the same for the private sector. As can be seen, if anything, the relative earnings in the government sector look more similar to the relative earnings under the counterfactual earnings schedule.

Based on a comparison of Panel A and Panel B, it might be tempting to conclude that moving to a private-sector based healthcare system might not bring the actual earnings any closer to the counterfactual earnings. However, the current private sector earnings in Turkey might not be the best approximation to the earnings in a healthcare system where the private sector is the largest player. The earnings in the private sector might have been largely shaped by the competition with the government sector in a country where the government employs more than 68% of the specialists.

Uncovering what the earnings across different specialties in Turkey would look like if the private sector became the dominant employer and how the resulting earnings would compare to the counterfactual earnings is well beyond the scope of this paper. Instead, I compare the counterfactual earnings to those in the US, where, in contrast to Turkey, the government employs only a small share—12%—of physicians (Popover et al. 2025). However, it is worth noting the caveat that even

in the US, the government plays an important role in doctors' earnings as discussed in Gottlieb et al. (2025), which makes this comparison far from perfect as well.

Panel C looks at the correlation between the relative earnings across specialties in the US and the counterfactual earnings schedule. As can be seen, there is virtually no correlation in the relative earnings. Specialties that receive a relatively low paycheck in the US, such as internal medicine and acute care, have higher relative earnings under the counterfactual income schedule. This might look surprising, given that the talent allocation (and hence, the relative demand for different specialties) in the US might be considered more "normal" based on the descriptive evidence in Section 3.5. The comparison in Panel C reveals that adopting an earnings schedule that looks similar to the one in a country where the talent allocation appears more intuitive in certain respects might not necessarily provide a solution.

6 Conclusions

I bring together data from different countries to study talent allocation across medical specialties and how it relates to earnings and working conditions. I combine original revealed-preference data from Turkey's residency match with surveys on subjective expectations, and estimate a structural model of specialty choice. Using the estimates from this model, I solve the problem of a planner that aims to fill all residency positions, and show that this earnings schedule also balances the distribution of talent across specialties, but has side-effects on gender composition.

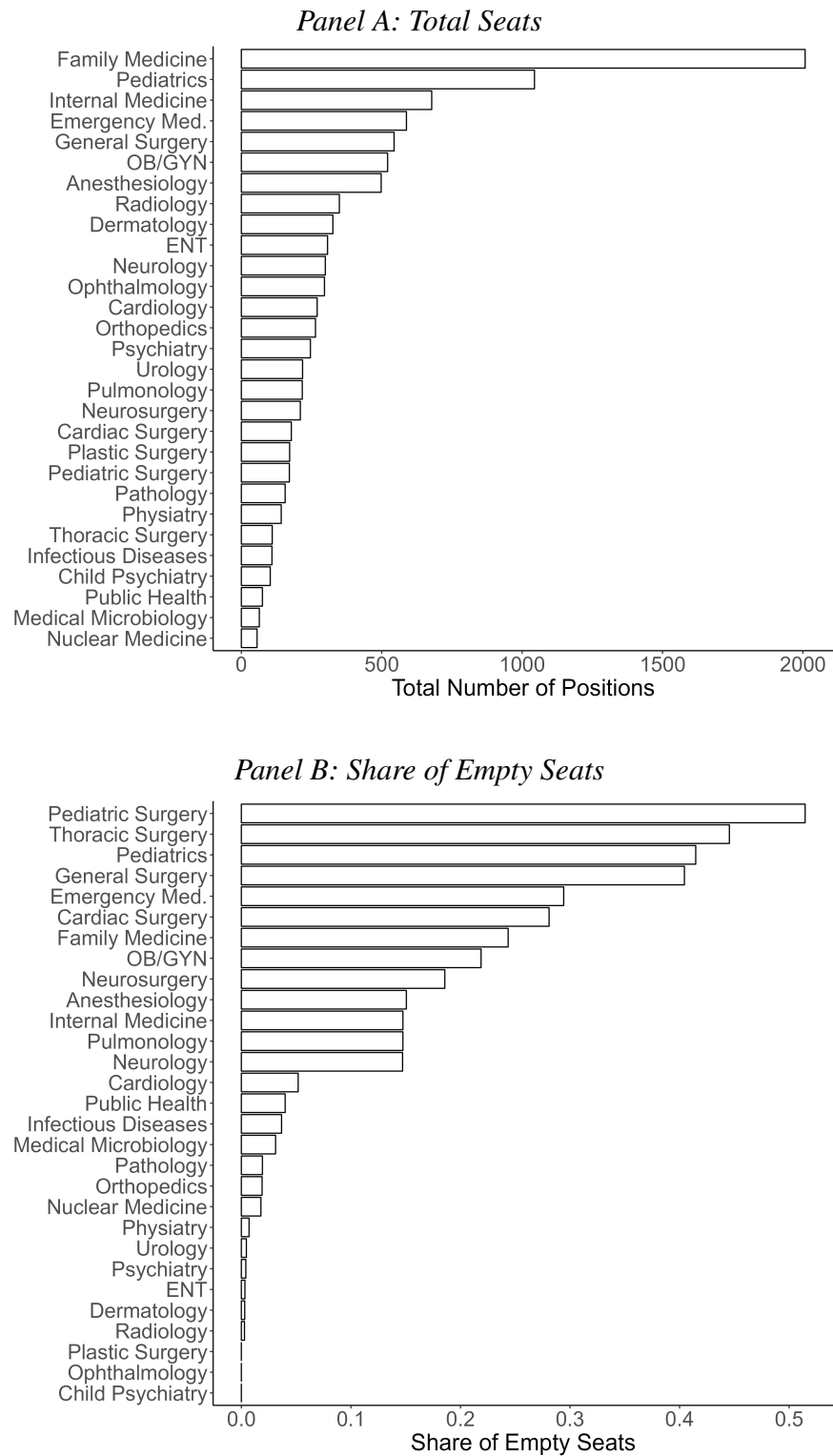
The counterfactual exercise in this paper is stylized and considers a planner who can set incomes. In reality, in many countries, the government and the private sectors co-exist as employers of the healthcare force. Studying the interactions between the two can allow answering how alternative earnings schedules can be achieved in practice.

References

- Attema, A. E., M. M. Galizzi, M. Groß, H. Hennig-Schmidt, Y. Karay, O. L’Haridon, and D. Wiesen (2023). The formation of physician altruism. *Journal of Health Economics* 87, 102716.
- Balinski, M. and T. Sönmez (1999). A tale of two mechanisms: Student placement. *Journal of Economic Theory* 84(1), 73–94.
- Berry, S., J. Levinsohn, and A. Pakes (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy* 112(1), 68–105.
- Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics* 25(2), 242–262.
- Bobba, M., T. Ederer, G. Leon-Ciliotta, C. Neilson, and M. G. Nieddu (2021). Teacher compensation and structural inequality: Evidence from centralized teacher school choice in Perú. NBER Working Paper 29068, National Bureau of Economic Research.
- Buehler, A., J. D. Gottlieb, J. Hicks, L. Laun, M. Palme, M. Polyakova, V. Udalova, and M. Ventura (2026, May). International comparison of physician incomes. *AEA Papers and Proceedings* 116.
- Carmody, B. (2020). How is the three-digit USMLE score calculated? The Sheriff of Sodium (blog). Accessed April 2026.
- Centre National de Gestion (2023). Arrêté du 10 octobre 2023 relatif à l’affectation des étudiants et des internes en médecine ayant satisfait aux épreuves classantes nationales anonymes. JORFTEXT000048264448. Available at <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000048264448>.
- Congressional Research Service (2025, December). Federal support for graduate medical education. CRS Report R48636, Congressional Research Service.
- Gottlieb, J. D., M. Polyakova, K. Rinz, H. Shiple, and V. Udalova (2025, May). The earnings and labor supply of U.S. physicians. *The Quarterly Journal of Economics* 140(2).
- Hatfield, J. W. and P. R. Milgrom (2005). Matching with contracts. *American Economic Review* 95(4), 913–935.
- Kelso, Jr., A. S. and V. P. Crawford (1982). Job matching, coalition formation, and gross substitutes. *Econometrica* 50(6), 1483–1504.
- Li, J. (2018). Plastic surgery or primary care? altruistic preferences and expected specialty choice of U.S. medical students. *Journal of Health Economics* 62, 45–59.
- Méndez, S. J., A. Scott, and P. Sivey (2025). Junior doctors’ specialty choice: Stated and revealed preferences over more than a decade. *Labour Economics* 94, 102697.

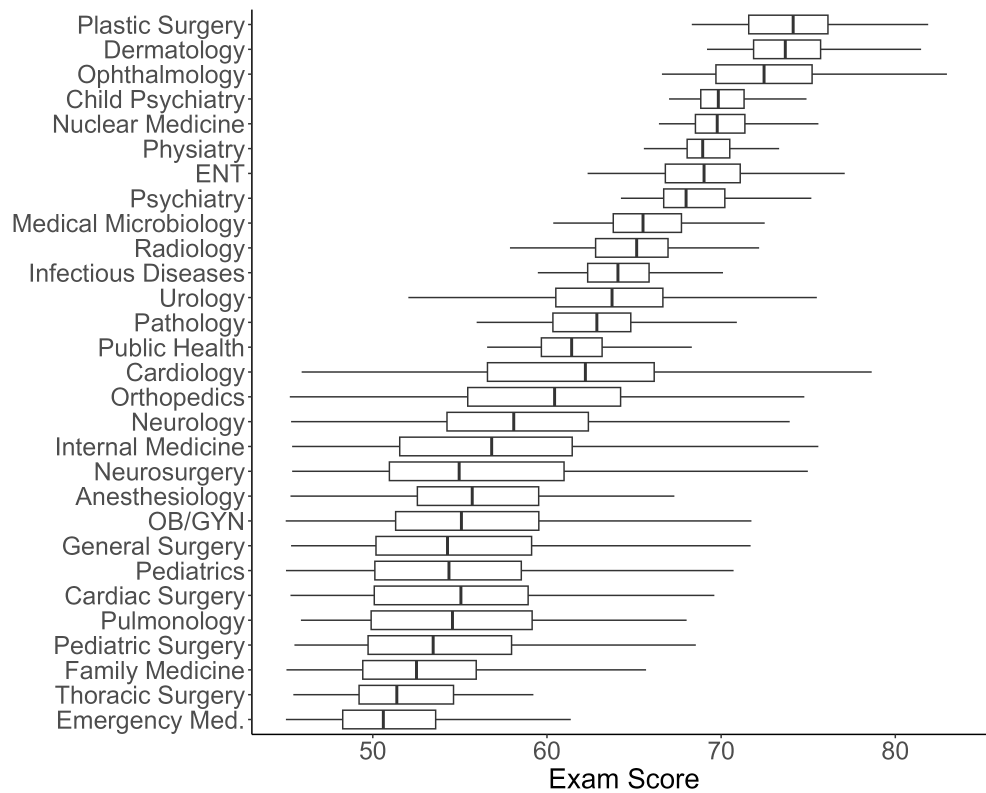
- National Resident Matching Program (2024). Charting outcomes in the match: Senior students of U.S. MD medical schools. Technical report, National Resident Matching Program, Washington, DC.
- Nicholson, S. (2002). Physician specialty choice under uncertainty. *Journal of Labor Economics* 20(4), 816–847.
- Norcini, J., I. Grabovsky, M. A. Barone, M. B. Anderson, R. S. Pandian, and A. J. Mechaber (2024). The associations between United States Medical Licensing Examination performance and outcomes of patient care. *Academic Medicine* 99(3), 325–330.
- NTV (2025, May). Doktorlar maliye’nin radarında. Accessed April 2026.
- OSYM (2025). Degerlendirme ve secme-yerlestirme islemleri. Accessed April 2026.
- Popover, J. L., T. Jones, C. Kalathia, A. Mackey, N. King, E. Sardzinski, Z. Oulton, A. Imam, M. Al-Masri, and P. G. Toomey (2025). Physician employment in America: Private practices dominate despite increased hospital employment. *JSLS: Journal of the Society of Laparoscopic & Robotic Surgeons* 29(2).
- Raymond, M. R., C. Stevens, and S. D. Bucak (2019). The optimal number of options for multiple-choice questions on high-stakes tests: Application of a revised index for detecting nonfunctional distractors. *Advances in Health Sciences Education* 24(1), 141–150.
- Sivey, P., A. Scott, J. Witt, C. Joyce, and J. Humphreys (2012). Junior doctors’ preferences for specialty choice. *Journal of Health Economics* 31(6), 813–823.
- the Official Gazette (2022). Tıpta ve dış hekimliğinde uzmanlık eğitimi yönetmeliği. Regulation on Specialty Training in Medicine and Dentistry.
- Turkish Statistical Institute (2023). Healthcare spending statistics in 2023. Accessed November 2025.
- Walters, C. R. (2024). Empirical Bayes methods in labor economics. NBER Working Paper 33091, National Bureau of Economic Research.
- Wasserman, M. (2023). Hours constraints, occupational choice, and gender: Evidence from medical residents. *Review of Economic Studies* 90(3), 1535–1568.

Figure 1: Capacity and Empty Seats Across Specialties



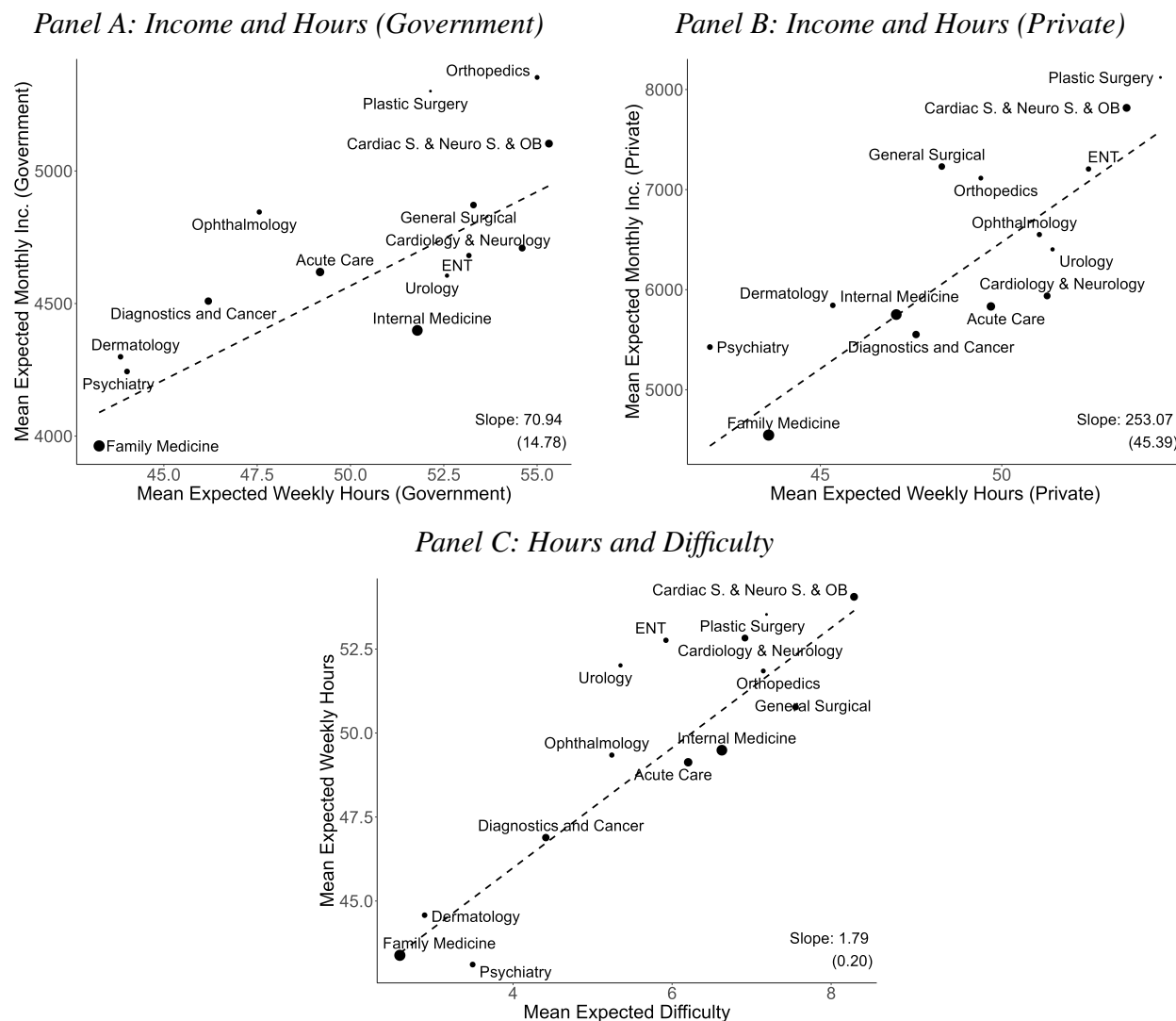
Notes: The figure shows the number of available seats (Panel A) and the share of empty seats (Panel B) by specialty in the Fall 2025 residency match in Turkey.

Figure 2: Distribution of Exam Scores by Specialty



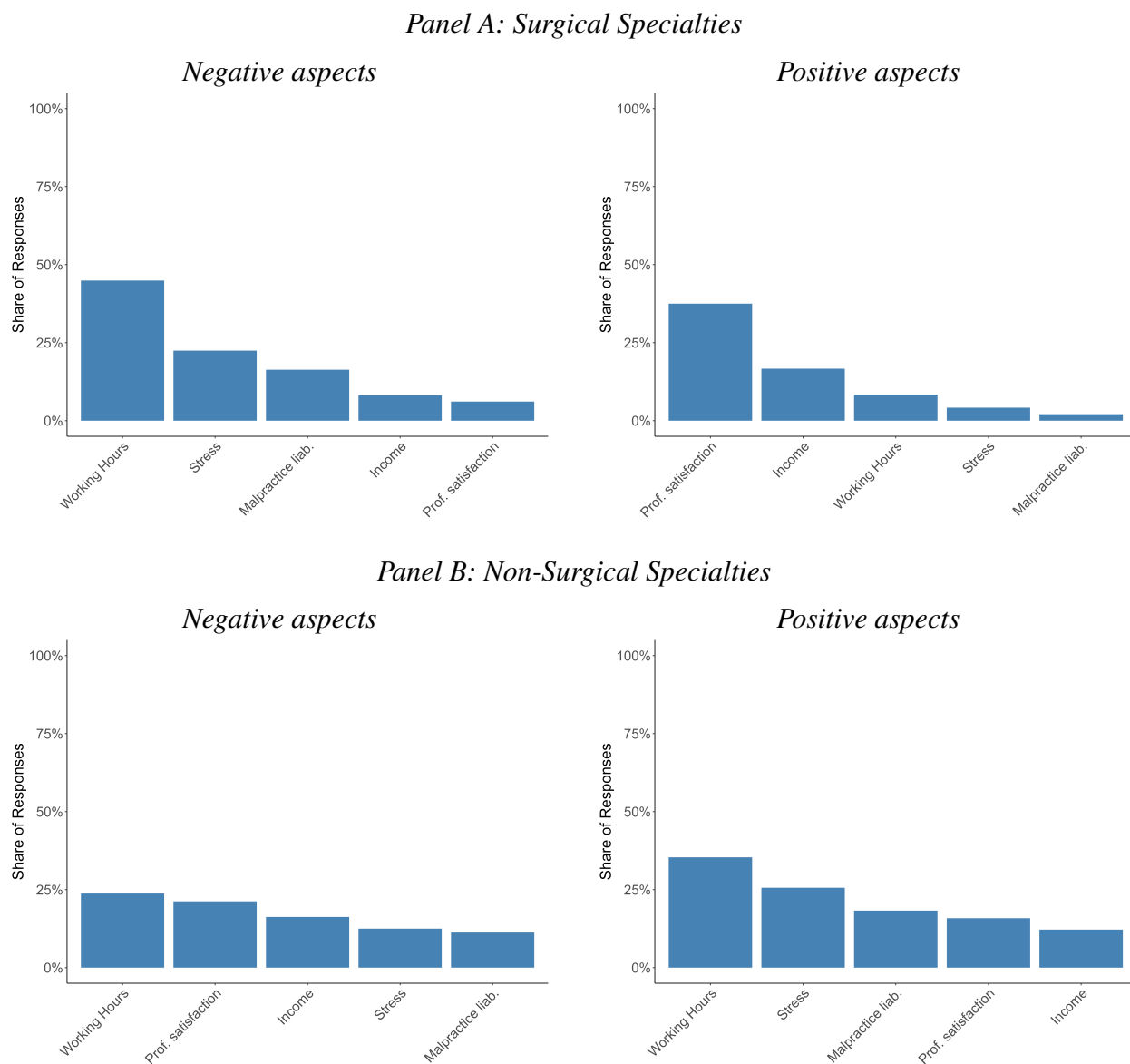
Notes: The figure shows the distribution of exam scores among the students who were predicted to be matched to given specialties on the last day of the match simulation.

Figure 3: Income, Working Hours, and Difficulty across Specialty Groups



Notes: The figure shows the relationships between survey measures of expected income, working hours, and difficulty across specialties. Panel A shows the relationship between expected income and working hours in the government sector. Panel B repeats the same for the private sector. Panel C shows the relationship between average working hours, constructed by taking a weighted average of working hours in both sectors where the weights are based on the average expected probability of working in the private sector, and the average difficulty score assigned to specialties by applicants. The dashed lines show the OLS regression fits. Slope coefficients with standard errors in parentheses are shown in the bottom right of each panel.

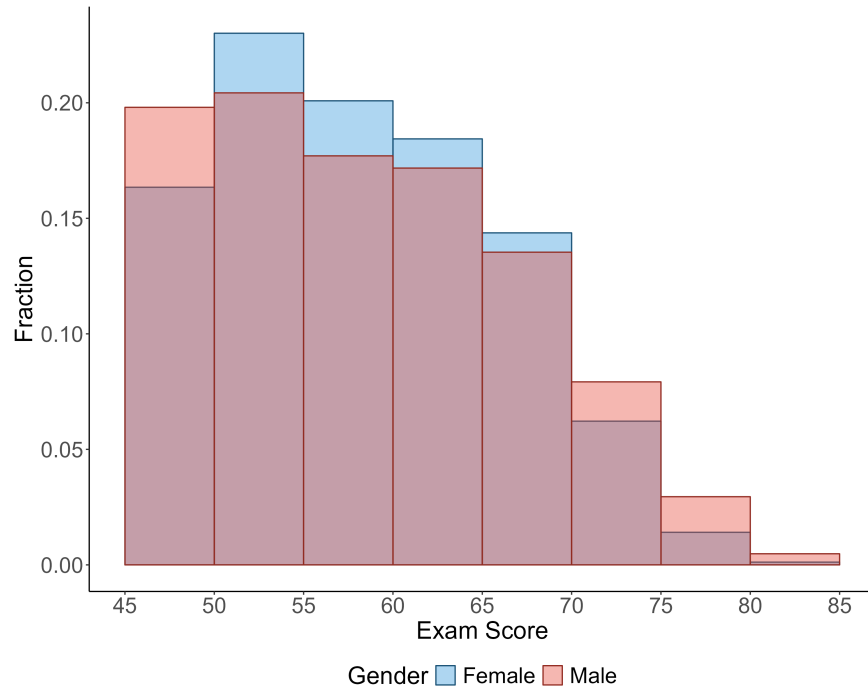
Figure 4: Survey on the positive and negative aspects, surgical vs non-surgical specialties



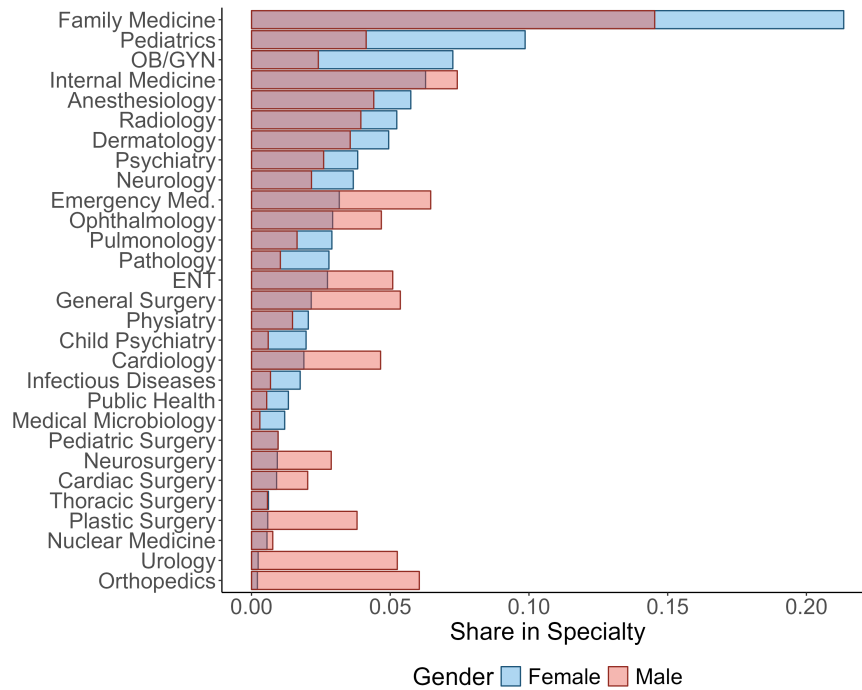
Notes: The figure shows the share of responses containing a keyword belonging to a given theme when the survey respondents were asked to state the negative (left plots) and positive (right plots) aspects of the specialty they chose. In Panel A, the plots show the shares of responses among survey respondents who plan to enroll in a surgical specialty. Panel B repeats the same structure for survey respondents who plan to enroll in a non-surgical specialty.

Figure 5: Exam Scores and Specialty Entry by Gender

Panel A: Exam Scores

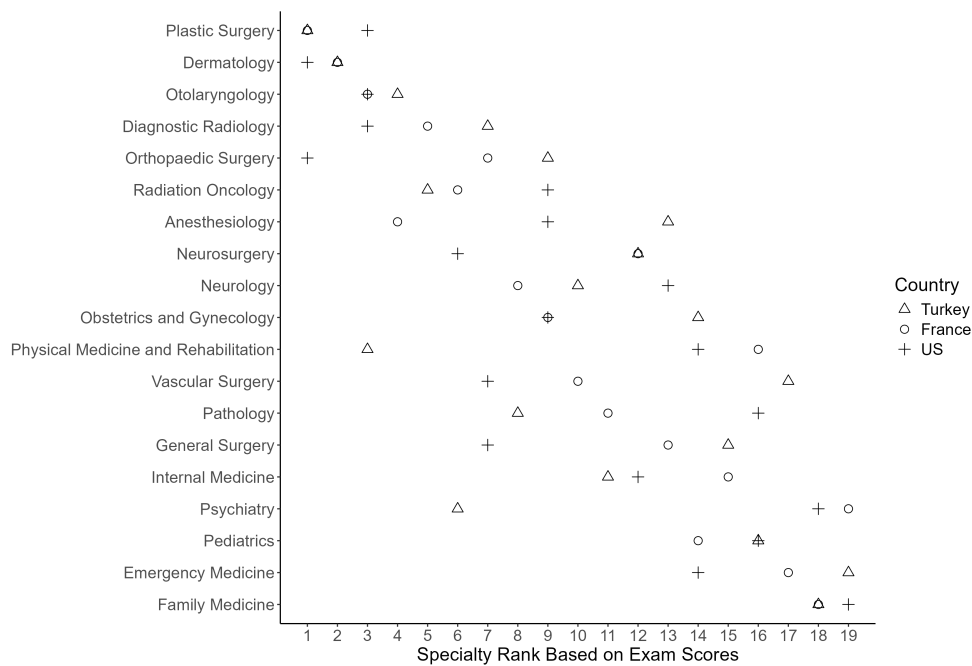


Panel B: Share of Matched Applicants



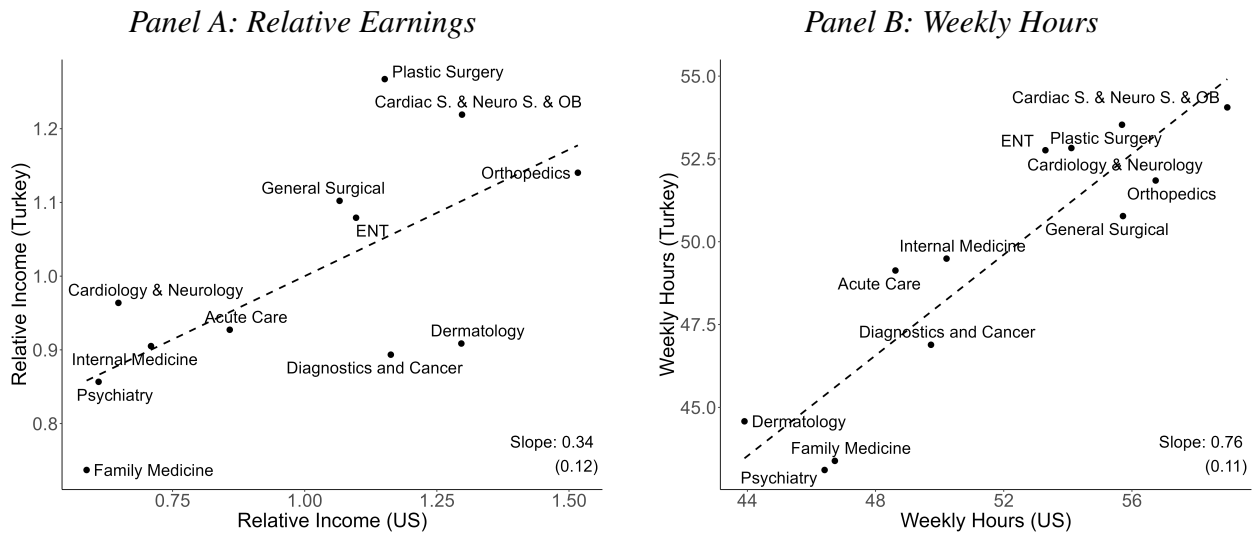
Notes: Panel A shows the distribution of exam scores by gender. Each bar represents the fraction of observations within each gender that fall into the specified score bin. Panel B shows the share of matched applicants in each specialty, separately by gender. Shares are computed within each gender.

Figure 6: Exam-Based Rankings of Specialties Across Countries



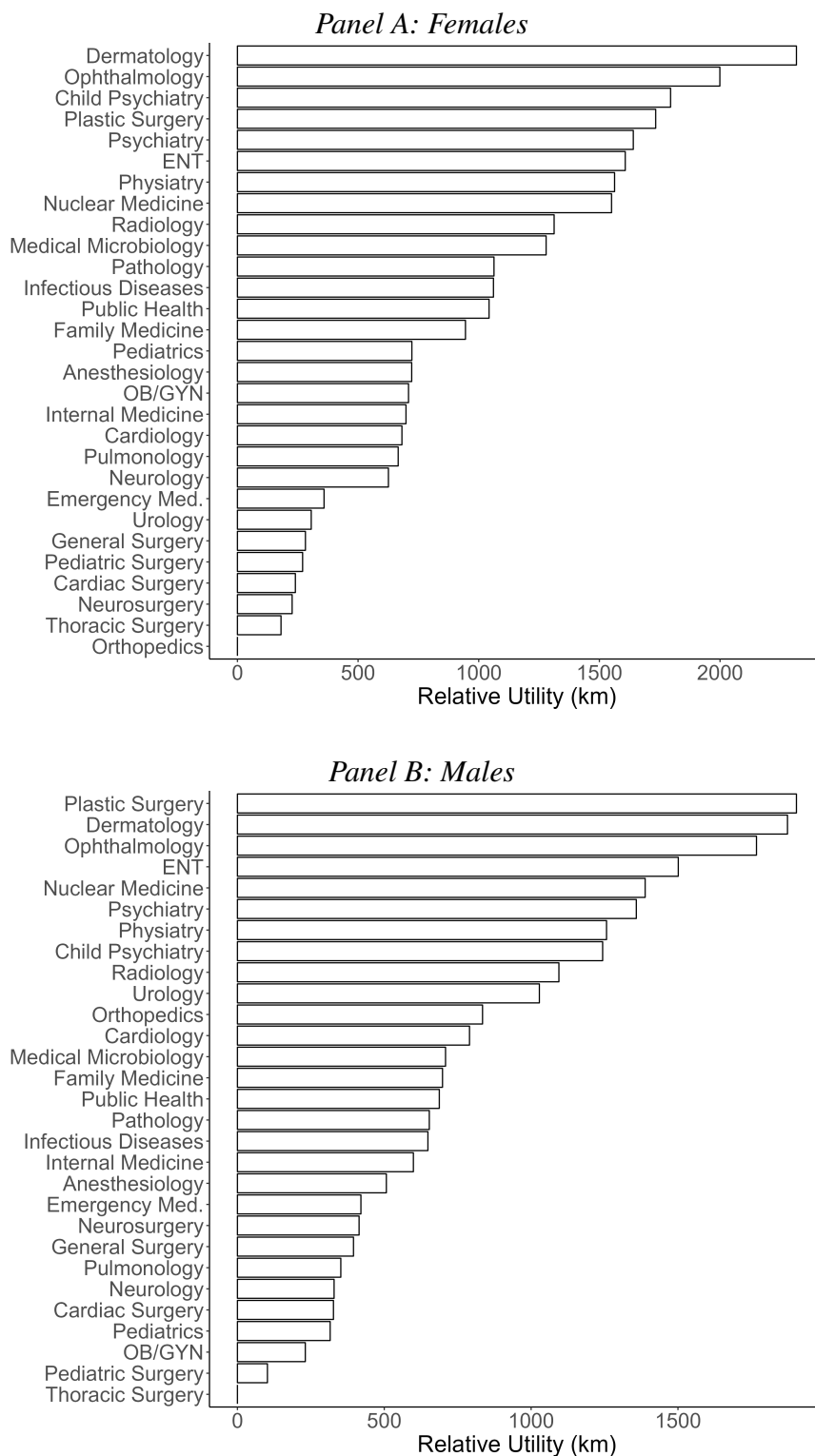
Notes: The figure shows the competitiveness of specialties across countries. US rankings are based on mean USMLE Step 2 CK scores of matched applicants (the specialty with the highest mean score gets the highest rank). France rankings are based on mean ECNi rank of matched students (the specialty with the lowest mean rank of ECNi scores gets the highest rank). Turkey rankings are based on mean TUS exam scores of matched applicants (the specialty with the highest mean score gets the highest rank). See Appendix Table 3 for the specialty mapping across countries. Specialties are sorted based on the average rank across countries.

Figure 7: Income and Working Hours across Specialties: US and Turkey



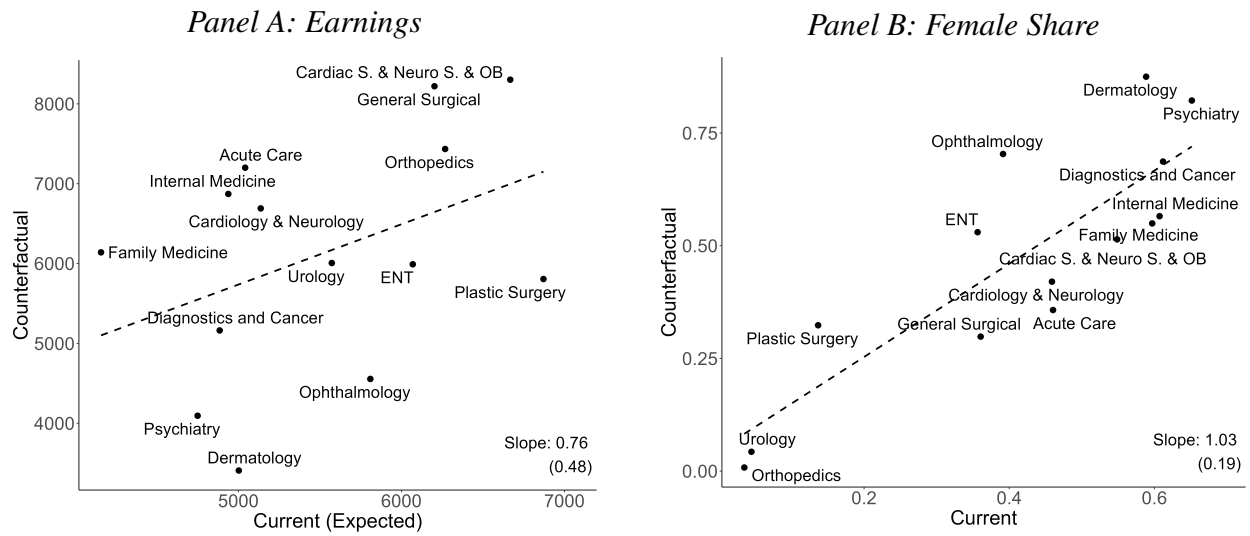
Notes: The figure compares income and working hours across specialty groups in the US and Turkey. Each point represents a specialty group described in Section 3.2. I map NRMP specialties in the US to these specialty groups. Turkish earnings and weekly hours are based on survey data and represent a weighted average of expected government and private sector earnings and working hours, where the weights are the expected probability of private sector employment as reported in the survey. US earnings and working hours are based on data from Gottlieb et al. (2025). When multiple NRMP specialties map to the same Turkish specialty group, I take the unweighted average across specialties. Panel A compares earnings, where earnings for each specialty are expressed relative to the cross-specialty group mean in each country. Panel B compares weekly hours. The dashed lines show the OLS regression fits. Slope coefficients with standard errors in parentheses are shown in the bottom right of each panel.

Figure 8: Relative Utility of Specialties by Gender



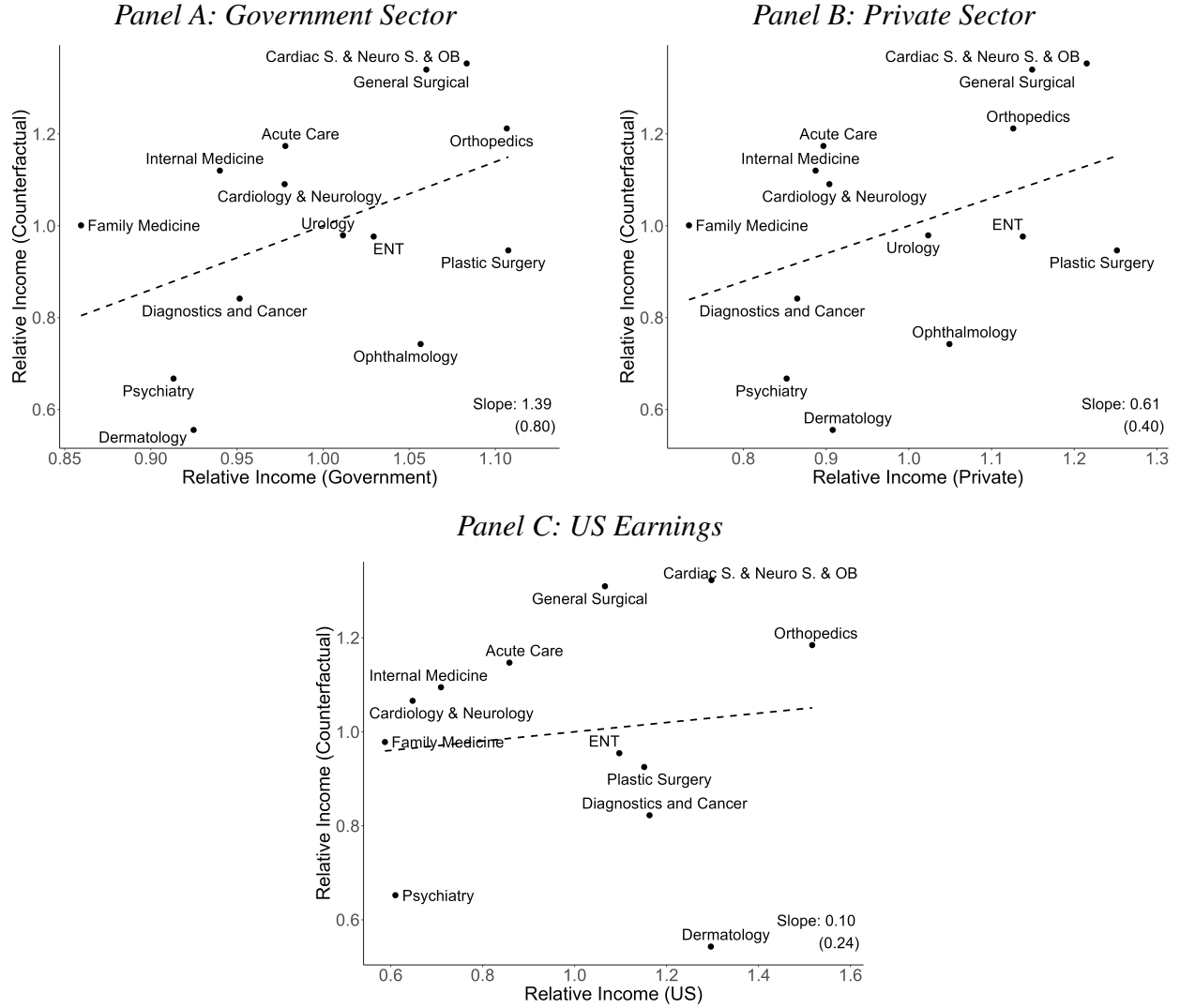
Notes: The figure shows the relative utility of each specialty, expressed in kilometers. For each gender, the mean utility for the specialty with the lowest utility is normalized to zero.

Figure 9: Counterfactual Earnings and Gender Composition



Notes: Panel A compares the average counterfactual monthly earnings for each specialty group with the current average earnings based on the surveys. Panel B compares the counterfactual female share in each specialty group with the current female share. In both panels, specialty-group level averages represent weighted averages across the specialties in the group, where the weights correspond to the number of doctors in each specialty under the counterfactual. The dashed lines show the OLS regression fits. Slope coefficients with standard errors in parentheses are shown in the bottom right.

Figure 10: Counterfactual and Current Relative Earnings Across Specialties



Notes: The figure compares the counterfactual earnings to the earnings in the government and the private sector in Turkey and to the overall earnings in the US. The comparisons are at the level of specialty groups as described in Appendix Table 1. To calculate specialty-group level counterfactual earnings, I take a weighted average of the counterfactual specialty earnings where the weights correspond to the number of filled seats in each specialty in the counterfactual scenario. In all panels, earnings are expressed relative to the cross-specialty group mean. In Panel A, I show the relationship between the counterfactual earnings and the expected earnings in the government sector. In Panel B, I show the relationship between the counterfactual earnings and the expected earnings in the private sector. In Panel C, I show the relationship between the counterfactual earnings and US earnings. US earnings are based on data from Gottlieb et al. (2025). US (NRMP) specialties are matched to Turkish specialty groups using the crosswalk in Appendix Table 4. When multiple NRMP specialties match to the same specialty group, an unweighted average of the earnings reported in Gottlieb et al. (2025) is used. The dashed lines show the OLS regression fits. Slope coefficients with standard errors in parentheses are shown in the bottom right of each panel.

Table 1: The association of performance in the USMLE and TUS

	β	SE
10 point increase in TUS score	0.070	(0.020)

Notes: The table shows the coefficient estimate and the standard error from an OLS regression of an indicator for whether a USMLE question, randomly chosen from a pool of translated questions, was answered correctly, on the TUS (Turkish medical specialty exam) score. The specification includes question fixed effects. The coefficient is rescaled to make it interpretable as the effect of a 10-point (one standard deviation) increase in TUS score.

Table 2: Coefficient Estimates by Gender

Variable	Females		Males	
	Coef.	S.E.	Coef.	S.E.
Anesthesiology	-5.7934	0.1107	-5.8484	0.1207
Cardiac Surgery	-7.3628	0.1939	-6.4091	0.1499
Cardiology	-5.9252	0.1466	-4.9673	0.1187
Child Psychiatry	-2.3013	0.1698	-3.5565	0.2464
Dermatology	-0.6034	0.1653	-1.5980	0.1588
ENT	-2.9125	0.1452	-2.7551	0.1294
Emergency Med.	-6.9758	0.1273	-6.1181	0.1130
Family Medicine	-5.0682	0.0956	-5.2533	0.1024
General Surgery	-7.2270	0.1424	-6.1968	0.1174
Infectious Diseases	-4.6907	0.1571	-5.4081	0.2242
Internal Medicine	-5.8706	0.1089	-5.5621	0.1097
Medical Microbiology	-3.9799	0.1863	-5.2213	0.3223
Neurology	-6.1075	0.1217	-6.4023	0.1446
Neurosurgery	-7.4064	0.1908	-6.1375	0.1346
Nuclear Medicine	-3.0976	0.2551	-3.1059	0.2296
OB/GYN	-5.8364	0.1075	-6.7072	0.1408
Ophthalmology	-1.6364	0.1559	-1.9265	0.1423
Orthopedics	-8.1446	0.3638	-4.8292	0.1130
Pathology	-4.6834	0.1355	-5.3934	0.1901
Pediatric Surgery	-7.2651	0.1897	-7.1086	0.1937
Pediatrics	-5.7912	0.1036	-6.4438	0.1240
Physiatry	-3.0567	0.1622	-3.5153	0.1775
Plastic Surgery	-2.5021	0.2593	-1.5031	0.1538
Psychiatry	-2.8042	0.1367	-3.1999	0.1524
Public Health	-4.7497	0.1739	-5.2875	0.2469
Pulmonology	-5.9757	0.1311	-6.3311	0.1594
Radiology	-3.8722	0.1181	-4.0200	0.1284
Thoracic Surgery	-7.5564	0.2274	-7.4267	0.2380
Urology	-7.1488	0.3448	-4.2270	0.1189
Distance (100km)	-0.3255	0.0097	-0.3112	0.0093
City Development Rank	-0.4898	0.0261	-0.3400	0.0248
University Hospital	-0.4413	0.0376	-0.4516	0.0379
City Hospital	0.4141	0.0416	0.4908	0.0425

Notes: The table shows the parameter estimates and standard errors from maximum likelihood estimation of equation 1. Development ranking refers to the ranking of cities based on an index published by the Ministry of Industry and Technology in 2019, with lower ranks corresponding to higher levels of development. I standardize this variable to have zero mean and unit variance. Distance is between the city where the doctor finished the medical school and the residency program.

Table 3: Second Stage Coefficient Estimates

Variable	Females		Males	
	Coef.	S.E.	Coef.	S.E.
Expected Income (100 USD)	0.130	0.060	0.160	0.060
Expected Weekly Hours	-0.290	0.120	-0.140	0.110
Difficulty	-0.550	0.270	-0.630	0.230

Notes: The table shows the coefficient estimates and standard errors from regressing mean specialty utility on specialty characteristics. Expected monthly income is calculated as a weighted average of mean income in the government and private sectors, where the weight on private sector income is the mean probability of working in the private sector reported for the specialty. Expected hours are calculated similarly. Means for all variables for all specialties are calculated using the procedure explained in Section 4.2.

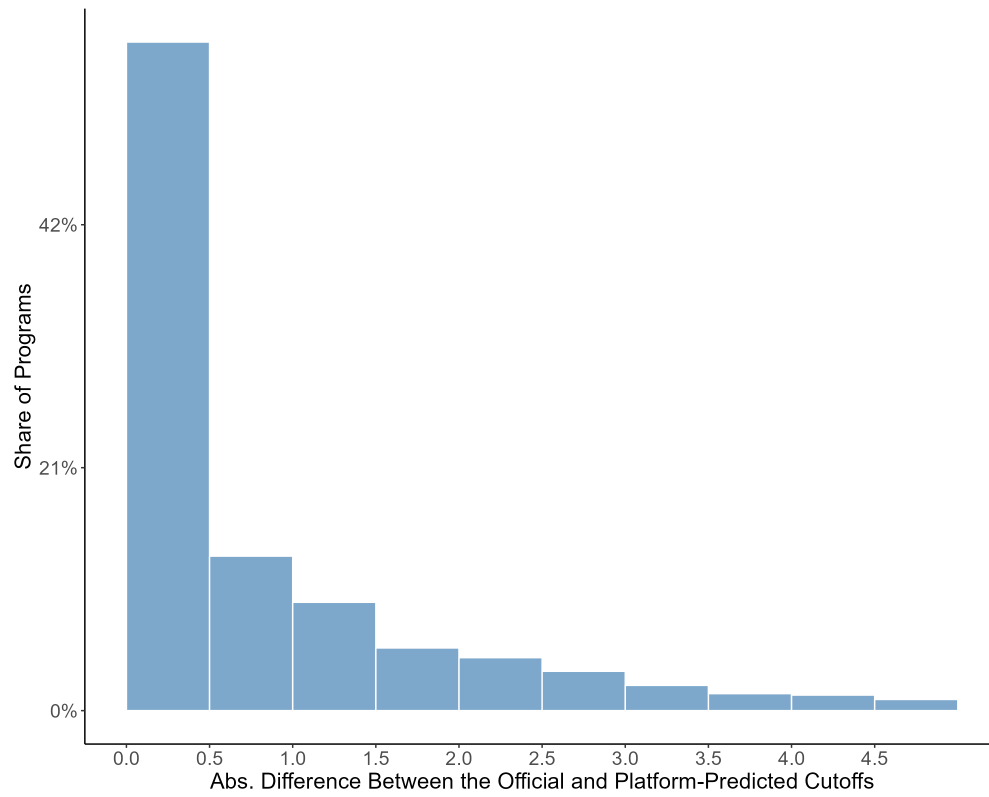
Table 4: Actual and Counterfactual Income, Fill Rates, Mean Scores, and Female Share by Specialty

Specialty	Actual				Counterfactual			
	Income (\$)	Fill (%)	Score	Women (%)	Income (\$)	Fill (%)	Score	Women (%)
Anesthesiology	5407.1	85.1	55.9	57.3	7618.9	100.0	66.0	48.3
Cardiac Surgery	7194.0	73.5	55.1	31.5	8629.4	100.0	56.8	30.2
Cardiology	5431.9	98.0	61.4	29.5	6480.1	100.0	60.0	30.0
Child Psychiatry	4352.5	100.0	70.4	77.1	3480.6	100.0	61.4	85.0
Dermatology	5002.6	100.0	74.0	58.9	3408.5	100.0	62.4	87.5
ENT	6069.6	100.0	69.0	35.6	5990.9	100.0	64.6	53.0
Emergency Med.	4635.4	64.3	51.4	33.5	6849.0	100.0	62.2	25.2
Family Medicine	4119.7	66.2	52.9	60.2	6303.7	100.0	61.9	53.3
General Surgery	6264.9	56.9	55.2	29.2	8308.6	100.0	59.9	21.9
Infectious Diseases	4185.3	100.0	64.2	72.5	4964.3	100.0	63.4	81.0
Internal Medicine	4977.8	79.8	57.3	46.5	7049.5	100.0	63.8	35.3
Medical Microbiology	4010.0	100.0	66.9	80.4	4370.9	100.0	62.7	88.3
Neurology	4819.2	82.1	58.4	63.6	6881.1	100.0	64.6	52.9
Neurosurgery	6726.9	86.3	56.4	25.0	8092.1	100.0	56.3	22.6
Nuclear Medicine	4790.0	100.0	69.9	42.9	3689.2	100.0	56.3	72.0
OB/GYN	6472.0	73.4	55.7	75.6	8276.6	100.0	58.9	69.0
Ophthalmology	5809.5	100.0	72.8	39.1	4555.3	100.0	60.9	70.3
Orthopedics	6268.1	96.8	60.0	3.5	7434.5	100.0	60.8	0.8
Pathology	4577.9	100.0	62.8	73.4	5261.1	100.0	60.3	77.0
Pediatric Surgery	6447.2	48.1	54.2	50.7	8292.6	100.0	56.4	42.5
Pediatrics	5187.5	52.8	55.2	71.1	7049.5	100.0	56.8	68.3
Physiatry	4516.3	100.0	69.2	58.8	3921.7	100.0	61.6	82.0
Plastic Surgery	6871.3	100.0	74.1	13.7	5806.4	100.0	59.2	32.4
Psychiatry	4913.7	100.0	68.6	60.3	4346.8	100.0	58.5	81.1
Public Health	4284.3	100.0	61.6	71.4	4723.7	100.0	57.4	75.7
Pulmonology	4495.5	88.6	54.7	64.5	6343.8	100.0	62.1	54.2
Radiology	5064.3	100.0	65.2	57.8	5581.9	100.0	62.6	62.4
Thoracic Surgery	5526.5	54.5	52.5	52.3	7651.0	100.0	58.5	50.5
Urology	5572.9	100.0	63.5	4.5	6006.9	100.0	58.4	4.3

Notes: The table shows the current and counterfactual monthly income, fill rates, and mean scores across specialties. The table excludes 13 specialties each of which has fewer than 0.5% of the available seats. The current mean income is computed using the Empirical Bayes procedure described in Section 4.2. The counterfactual income is computed by solving the optimization problem of filling all residency positions while minimizing the total payments to physicians, as outlined in Section 5.

A Appendix

Appendix Figure 1: Platform-Predicted vs. Realized Cutoff Scores



Notes: The figure shows the distribution of differences between platform-predicted and realized cutoff scores across residency programs in the Fall 2025 match.

Appendix Table 1: Specialty Groupings and the Required Years of Training

Specialty	Group	Years of Training
Anesthesiology	Acute Care	5
Emergency Med.	Acute Care	4
Cardiac Surgery	Cardiac S. & Neuro S. & OB	5
Neurosurgery	Cardiac S. & Neuro S. & OB	5
OB/GYN	Cardiac S. & Neuro S. & OB	4
Cardiology	Cardiology & Neurology	5
Neurology	Cardiology & Neurology	5
Dermatology	Dermatology	4
Medical Genetics	Diagnostics and Cancer	4
Nuclear Medicine	Diagnostics and Cancer	4
Pathology	Diagnostics and Cancer	4
Radiation Oncology	Diagnostics and Cancer	5
Radiology	Diagnostics and Cancer	4
ENT	ENT	5
Family Medicine	Family Medicine	3
Physiatry	Family Medicine	4
Sports Medicine	Family Medicine	4
General Surgery	General Surgical	5
Pediatric Surgery	General Surgical	5
Thoracic Surgery	General Surgical	5
Infectious Diseases	Internal Medicine	4
Internal Medicine	Internal Medicine	4
Pediatrics	Internal Medicine	4
Pulmonology	Internal Medicine	4
Ophthalmology	Ophthalmology	4
Orthopedics	Orthopedics	5
Plastic Surgery	Plastic Surgery	5
Child Psychiatry	Psychiatry	4
Psychiatry	Psychiatry	4
Urology	Urology	5
Aerospace Medicine	Science & Other	3
Anatomy	Science & Other	3
Forensic Medicine	Science & Other	4
Histology and Embryology	Science & Other	4
Medical Biochemistry	Science & Other	4
Medical Ecology and Hydroclimatology	Science & Other	3
Medical Microbiology	Science & Other	4
Medical Pharmacology	Science & Other	4
Physiology	Science & Other	4
Public Health	Science & Other	4
Underwater and Hyperbaric Medicine	Science & Other	3

Notes: The table shows the list of specialty programs in Turkey, the specialty grouping used in the descriptive analysis in Section 3, and the required years of training for each specialty.

Appendix Table 2: Relationship Between Specialty Group Difficulty and Open-Ended Response Themes

<i>Panel A: Positive Aspects</i>					
	Stress	Malpractice	Working Hours	Income	Prof. Satisfaction
Difficulty	-0.056 (0.017)	-0.055 (0.015)	-0.066 (0.020)	0.006 (0.016)	0.067 (0.019)
<i>Panel B: Negative Aspects</i>					
	Stress	Malpractice	Working Hours	Income	Prof. Satisfaction
Difficulty	0.032 (0.017)	0.031 (0.016)	0.065 (0.021)	-0.006 (0.016)	-0.039 (0.017)

Notes: Each column reports the coefficient from an OLS regression of a binary indicator for whether the response to a question about the positive (or negative) aspects of a specialty contains a keyword related to a topic on group-level average expected difficulty of the specialty chosen by the respondent. Standard errors are in parentheses. Panel A uses responses to the question about positive aspects of a specialty. Panel B uses responses to the question about negative aspects.

Appendix Table 3: Specialty Mapping Across Countries

US	Turkey	France
Anesthesiology	Anesthesiology and Reanimation	Anesthesiology and Critical Care
Dermatology	Dermatology	Dermatology
Diagnostic Radiology	Radiology	Radiology and Medical Imaging
Emergency Medicine	Emergency Medicine	Emergency Medicine
Family Medicine	Family Medicine	General Medicine
General Surgery	General Surgery	Visceral and Digestive Surgery
Internal Medicine	Internal Medicine	Internal Medicine and Clinical Immunology
Neurology	Neurology	Neurology
Neurosurgery	Neurosurgery	Neurosurgery
Obstetrics and Gynecology	Obstetrics and Gynecology	Obstetrics and Gynecology
Orthopaedic Surgery	Orthopedics and Traumatology	Orthopedic Surgery
Otolaryngology	Otorhinolaryngology (ENT)	ENT and Cervico-Facial Surgery
Pathology	Medical Pathology	Pathology
Pediatrics	Pediatrics	Pediatrics
Physical Medicine and Rehabilitation	Physical Medicine and Rehabilitation	Physical Medicine and Rehabilitation
Plastic Surgery	Plastic Surgery	Plastic Surgery
Psychiatry	Psychiatry	Psychiatry
Radiation Oncology	Radiation Oncology	Oncology
Vascular Surgery	Cardiovascular Surgery	Vascular Surgery
<i>Unmapped</i>		
–	Aerospace Medicine, Anatomy, Cardiology, Child and Adolescent Psychiatry, Forensic Medicine, Histology and Embryology, Infectious Diseases, Medical Biochemistry, Medical Ecology and Hydroclimatology, Medical Genetics, Medical Microbiology, Medical Pharmacology, Military Health Services, Nuclear Medicine, Ophthalmology, Pediatric Surgery, Physiology, Public Health, Pulmonology, Sports Medicine, Thoracic Surgery, Underwater and Hyperbaric Medicine, Urology	Allergology, Cardiology, Endocrinology Diabetes and Metabolic Diseases, Forensic Medicine and Medical Expertise, Gastroenterology and Hepatology, Geriatrics, Hematology, Infectious Diseases, Intensive Care Medicine, Maxillofacial Surgery, Medical Biology, Medical Genetics, Medical Gynecology, Nephrology, Nuclear Medicine, Occupational Medicine and Health, Ophthalmology, Oral Surgery, Pediatric Surgery, Public Health, Pulmonology, Rheumatology, Thoracic and Cardiovascular Surgery, Urology, Vascular Medicine

Notes: This table shows the specialty mapping used for the cross-country comparison.

Appendix Table 4: Mapping of NRMP Specialties to Turkish Specialty Groups

NRMP Specialty	Turkish Specialty Group
Anesthesiology	Acute Care
Emergency Medicine	Acute Care
Neurological Surgery	Cardiac S. & Neuro S. & OB
ObGyn	Cardiac S. & Neuro S. & OB
Neurology	Cardiology & Neurology
Dermatology	Dermatology
Pathology	Diagnostics and Cancer
Radiation Oncology	Diagnostics and Cancer
Radiology	Diagnostics and Cancer
Otolaryngology	ENT
Family Medicine	Family Medicine
Physical Medicine and Rehabilitation	Family Medicine
Surgery	General Surgical
Internal Medicine	Internal Medicine
Pediatrics	Internal Medicine
Orthopaedic Surgery	Orthopedics
Plastic Surgery	Plastic Surgery
Psychiatry	Psychiatry

Notes: This table shows the mapping between NRMP specialties and the Turkish specialty groups used in Figure 7. Turkish specialty groups are as described in Section 3.2.