Gender Differences in Early Occupational Choices: Evidence from Medical Specialty Selection

Josep Amer-Mestre
Noémi Berlin
Agnès Charpin
Magali Dumontet

2024-5 Document de Travail/ Working Paper

EconomiX - UMR 7235 Bâtiment Maurice Allais
Université Paris Nanterre 200, Avenue de la République
92001 Nanterre Cedex

Site Web : economix.fr
Contact : secreteriat@economix.fr
Twitter : @EconomixU

Gender Differences in Early Occupational Choices: Evidence from Medical Specialty Selection

Josep Amer-Mestre[*,†], Agnès Charpin[‡,✉], Noémi Berlin[§], Magali Dumontet[¶]

January, 2024

Abstract

This paper analyses gender differences in occupational choices in a setting in which observed matches are solely determined by supply-side factors: the French centralised medical residency selection mechanism. We show that men and women facing the same occupational choice set make drastically different occupational choices. Medical specialties selected by women pay less, have lower time requirements, and are less competitive. To understand these differences and estimate how much of the gender gap in specialty sorting can be explained by individual preferences for job attributes, we administer a survey to prospective medical residents just before their specialty choice. Using both a hypothetical job choice framework and stated preferences, we show that while "hard" job characteristics (earnings, time requirements) only slightly reduce the gender gap in sorting, "soft" characteristics (daily tasks, contact with patients, willingness to help others) play a larger role in reducing the gap. We also find suggestive evidence of an anticipation effect of fertility on women's career choices. Our results suggest that individual preferences play a determinant role in explaining gender-based occupational segregation.

Keywords: Occupational segregation, Gender, Labour market, Job attributes, Willingness to pay

JEL Classification: J16, J22, J24, J31

[*] Department of Economics, European University Institute, Italy, josep.amer@eui.eu.

[†] European Commission, Joint Research Centre (JRC), Ispra, Italy.

[‡] Corresponding author: EconomiX, Université Paris Nanterre, France. agnes.charpin@parisnanterre.fr

[§] CNRS, EconomiX, Université Paris Nanterre, France. noemi.berlin@parisnanterre.fr

[¶] Université Paris Nanterre, EconomiX, CNRS, France. magali.dumontet@parisnanterre.fr

An earlier version of this paper appeared as chapters in Amer-Mestre's and Charpin's respective PhD theses and was circulated as an EUI Working Paper under the same title.

We are thankful to Alícia Adserà, Manuel Bagues, Michèle Belot, Francine Blau, Anne Boring, Karol Borowiecki, Joan Costa-i-Font, Thomas Crossley, Christian Dustmann, Alessandro Ferrari, Andrea Ichino, Nagore Iriberri, Essi Kujansuu, Melanie Lührmann, Karen Macours, Gregor Pfeifer, Hillel Rapoport, Anna Raute, Mirjam Reutter, Evan Riehl, Magali Schmidt, Arthur Schram, Alessandro Tarozzi, Alessandro Tondini, seminar participants at the 2021 ESPE annual conference, 2022 Dulbea Workshop in Gender Economics, 2022 Royal Economic Society annual conference, 2022 Society of Labor Economists annual conference, 2022 Spring Meeting of Young Economists, 2022 Society of Economics of the Household annual conference, 2022 French Economic Association annual conference, Labor Work in Progress Seminar at Cornell University, European Association of Labour Economics 2022 annual meeting, 2022 European Economic Association annual conference, 17th Belgian Day for Labour Economists, and EUI Microeconometrics Working Group for helpful comments. We are very grateful to the French National Research Agency for funding the MEDSPE project through an ANR JCJC grant (N°ANR-21-CE26-0013-01), and to the EUI for funding the lottery prizes through the Early Stage Research grant. We acknowledge support from UNESS (Université numérique en santé et sport) and specifically Olivier Palombi for providing a means to distribute our surveys, and from the DREES (Direction de la recherche, des études, de l'évaluation et des statistiques), the French Ministry of Health and ADISP (Archives des données issues de la statistique publique) for producing and distributing the INSEE-CNAMTS-DGFiP dataset on the earnings of private physicians. We also thank the Conférence des doyens de médecine and ANEMF for the support, and the Fédération S2CH - CNRS - UP1 - UPN - PSE for their technical support on the first survey. This project received IRB approval from Université Paris Nanterre for the first wave of the survey (Approval Number 2022-03-03), and from the EUI and ULB for the follow-up survey. All remaining errors are our own.

1 Introduction

In recent decades, women have made important progress in the labour markets of most developed countries, resulting in what Goldin (2014) refers to as one of the "grandest advances in society and the economy": the converging roles of men and women. Women have overtaken men in educational attainment, and although gender gaps in labour force participation and earnings remain sizable, they have narrowed. Despite this convergence, gender-based occupational segregation remains a strong feature of labour markets around the world and is the most important measurable factor explaining the gender wage gap (Blau and Kahn, 2017).

The reasons behind gender-based occupational segregation are difficult to identify from realised labour market outcomes, which result from a variety of factors such as worker productivity, search and matching frictions, and differences in the set of jobs available to workers. Such confounding factors hinder the identification of the different channels that contribute to gender-based occupational segregation.

In this paper, we address this fundamental identification problem by analysing gender differences in early occupational choices in a context where, by design, the demand side of the job market plays no role in determining the observed matches. Observing jobseekers' decisions as well as their occupational choice sets allows us to focus entirely on the role of supply-side factors in determining gender-based occupational segregation. It makes this paper the first one to precisely measure the magnitude of worker sorting in a real-world setting. In particular, we exploit the French centralised allocation mechanism of medical students to residency positions and analyse the occupational choices of these highly-skilled individuals using both administrative and collected survey data.

Each year, French medical students at the end of their sixth year of studies take a national exam called the National Ranking Examinations (*Épreuves classantes nationales* in French, hereafter NRE) after which they are ranked solely according to their performance in the exam. Students' rank determines the order in which they choose their residency position, defined as a medical specialty-location pair. Vacancies are filled on a first-ranked, first-served basis until all candidates have been allocated to a position, following a serial dictatorship mechanism.[1]

This mechanism offers three unique features to study gender differences in occupational choices. First, unlike in most job markets, employer preferences play no role in the matching process: employers do not screen, evaluate, or decide whether to hire candidates. The decision-making power is thus entirely on the candidates' side. As a result, there is no room for immediate hiring discrimination, nor for bargaining over working conditions, such as wages.[2]

[1] Throughout, a *position* is a (medical specialty, location) pair, while a *vacancy* is one of the available slots offered within a position. Each year, the Ministry of Health publishes a list of available vacancies. Note that the number of vacancies is always larger than the number of students to be allocated.

[2] Even though immediate hiring discrimination is ruled out, students' expectations of future discrimination by employers and patients may play a role in specialty selection. We discuss how such expectations may affect differential gender sorting

Second, the allocation mechanism we rely on is incentive-compatible: the candidates' optimal strategy is always to select their most preferred position in their choice set. Third, there is perfect information about each individual's choice set, which we can observe. This considerably minimises the potential role of gender differences in the ability to search and apply for jobs in explaining gender differences in occupational choices.

These three features imply that, unlike in most labour market settings, we are able to observe each jobseeker's optimal occupational choice, while abstracting from the demand side of the market and therefore focusing entirely on the supply side. After conditioning on individual choice sets, gender differences in occupational choices are entirely attributable to supply-side factors and are thus largely determined by preferences.[3] Whether the observed individual occupational choices come from intrinsic tastes, expectations of future discrimination, gender norms, or stereotypes, we label them as *preferences* as long as they are unconstrained by position availability.

In addition to these three unique features, this setting allows us to identify a group of individuals who, given their performance in the national exam, make their occupational choice when *all* positions still have at least one vacancy and thus choose their preferred medical specialty in their preferred location. This feature has one major implication: unlike in most settings in which the researcher only observes realised outcomes, we are certain to observe the most preferred occupational choice of each of these individuals who are unconstrained by position availability.

Furthermore, the specialisation choice that we focus on is decisive for medical students, as it determines the field in which they will specialise and work for the rest of their careers, which in turn has a major impact on their future earnings. Finally, jobseekers in this market form a very homogeneous group of young and highly skilled individuals holding the same formal qualifications, which reduces the existence of potential factors confounding the occupational decision, such as childbearing, even further.

We start our analysis by providing evidence on the importance of specialty choice in explaining the gender pay gap in the medical profession. We use individual-level administrative data containing information on the activity and billing of all self-employed physicians, who account for more than half of all physicians in France. We focus on the regulated-fee sector in which physicians' pay is not tied to tenure, it is not subject to convex returns to hours worked, and cannot be negotiated. We uncover that female physicians established in the regulated-fee sector earn about 9 percent less for each procedure they perform than their male counterparts. Conditioning on individual and practice characteristics does not substantially reduce the gender pay gap per procedure. However, we show that medical specialty sorting

across specialties in the results section.

[3]Note that it is beyond the scope of this paper to pin down the mechanisms leading to the formation of preferences for occupations. Women might be less willing to self-select into "male occupations" because it is not what is socially expected of them (Akerlof and Kranton, 2000), or they might refrain from entering inflexible occupations because they have internalised the fact that they will be expected to juggle family and work (Goldin, 2004, 2006).

alone reduces the residual gender pay gap per procedure by 43 percent (from 9 to 4 percent).[4] These results provide clear evidence of the importance of specialty choice for the labour market outcomes of medical students, and thus lay the first stone in motivating our analysis of gender-based occupational segregation in the medical context.

We then turn to the core of the paper and document the gender-based occupational segregation that exists among French medical students. Using unique individual-level administrative data on the French centralised allocation mechanism of medical students to residency positions, we find that conditional on facing the same choice set, men and women make drastically different occupational choices. Strikingly, we find that this is true even among top exam performers, who face no external constraints on their choices. This novel result implies that men and women facing the same choice set *prefer* different occupations. This is all the more interesting as one might expect gender differences in labour market outcomes to be limited in a group of such highly-educated individuals.

Next, we show that conditional on facing the same occupational choice set, women are more likely than men to choose specialties that have lower expected earnings and time requirements. They are also more likely than men to self-select into less competitive environments, as reported by workers in these occupations. To better understand the drivers of these sorting differentials, we turn to the field and survey a sample of exam takers five weeks before the occurrence of the 2022 national exam. We ask these medical students about their occupational aspirations, their expected performance in the exam, the drivers of their choices, and relevant demographic characteristics.

Using a hypothetical job choice framework, we show that taken as homogeneous groups, the men and women of our sample have remarkably similar willingness to pay (WTP) for different job attributes. When looking at the tails of the WTP distributions, our results do point towards the existence of a group of women with a particularly strong taste for time flexibility, who are willing to give up important shares of their earnings to avoid working longer, more inconvenient hours. Nonetheless, we do not find evidence of strong gender differences in WTP that could fully explain the observed gender differences in specialty sorting. Turning to a series of complementary questions eliciting self-reported preferences for job attributes, we additionally show that women care significantly more than men about having a job that implies patient care (contact with patients and willingness to help others), and that men put more importance than women on the prestige and earnings levels of their job.

Finally, we estimate how much of the gender gap in specialty sorting can be explained by these estimated individual-level preferences for job attributes. We show that while "hard" job characteristics such as earnings and time constraints only slightly reduce the gender gap in sorting, accounting for "soft" job characteristics such as daily tasks, contact with patients, and willingness to help others substantially

[4] Results are similar when focusing on the so-called free-billing sector in which the gender pay gap is 26 percent and occupational sorting alone explains 66 percent of it, leaving the gender pay gap per procedure at 9 percent.

reduces the sorting gap. We run additional analyses that examine the role of preferences for future fertility. Although the estimates are not statistically robust, we find evidence that the desire to have children in the near future is a relevant factor in explaining differences in specialty sorting among women but not among men, uncovering an anticipation effect of fertility on the career choices of women.

We first and foremost contribute to the literature on gender-based occupational segregation. To the best of our knowledge, ours is the first analysis relying on revealed preferences for jobs that provides evidence of gender-based occupational segregation in a labour market where sorting comes exclusively from the supply side. Our paper thus improves on existing work in which the role played by demand-side factors and differences in the jobseekers' formal qualifications cannot be fully accounted for when identifying gender-based occupational segregation (Altonji and Blank, 1999; Cortes and Pan, 2018; Lordan and Pischke, 2022).[5] It also complements Wasserman (2022), who shows in the U.S. context that reducing the hours requirements of a medical specialty boosts women's entry into that specialty but does not affect men's. While she provides evidence that the increase in women's entry is due to changes in labour supply rather than labour demand, we completely abstract from demand-side forces and show that supply-side decisions largely drive sorting into medical specialties.

Second, we take part in the current discussion of new classes of explanations for gender differences in labour market outcomes.[6] These include differences in preferences for certain workplace amenities (e.g. Sasser, 2005; Fortin, 2008; Cortes and Pan, 2018; Azmat and Ferrer, 2017; Wiswall and Zafar, 2018; Le Barbanchon, Rathelot and Roulet, 2020; Fluchtmann et al., 2020; Fadlon, Lyngse and Nielsen, 2020; Cook et al., 2021; Wasserman, 2022; Bolotnyy and Emanuel, 2022; Lordan and Pischke, 2022) differences in personality traits (e.g. DeLeire and Levy, 2004; Dohmen et al., 2011; Flory, Leibbrandt and List, 2014; Buser, Niederle and Oosterbeek, 2014; Reuben, Sapienza and Zingales, 2015; Reuben, Wiswall and Zafar, 2017; Buser and Yuan, 2019; Biasi and Sarsons, 2020; Díez-Rituerto et al., 2022), and social norms about what women can and should do (e.g. Akerlof and Kranton, 2000; Goldin, 2002; Charles, Guryan and Pan, 2018; Porter and Serra, 2020). In addition, our survey results are in line both with the series of papers by Goldin and co-authors arguing that the gender pay gap would be greatly reduced if workplaces did not reward individuals who work long, or particular hours (Bertrand, Goldin and Katz, 2010; Goldin, 2014, 2015; Goldin and Katz, 2016) and those highlighting the costs of fertility incurred by women well before the birth of their first child (Adda, Dustmann and Stevens, 2017; Wasserman, 2023).

Third, we contribute to the literature on the determinants of medical specialty choice. One of the

[5] Demand-side factors that have been identified by the literature as the main drivers of gender-based occupational segregation are discrimination and human capital accumulation (Altonji and Blank, 1999). The former channel has received support from a growing empirical literature documenting the existence of a gap in the probability that men and women are interviewed or hired for the same job (e.g. Goldin and Rouse, 2000; Riach and Rich, 2002; Rich, 2014; Neumark, 2018). Regarding the latter channel, the literature shows that although women have now overtaken men in terms of educational attainment (Goldin, Katz and Kuziemko, 2006), there still exists marked gender differences in labour force participation and career development (Bertrand, Goldin and Katz, 2010).

[6] See Bertrand (2011), Cortes and Pan (2018) and Azmat and Petrongolo (2014) for detailed accounts of the directions that the literature has taken in recent years.

closest papers to ours in this literature is Sivey et al. (2012), both in terms of the methodology used to elicit individual tastes for job attributes and of the variety of job attributes of interest. Nonetheless, our setting allows us to go one step further and document that, even in a setting in which only supply-side factors are at play, there are large gender differences in specialty choices. In this respect, we more generally improve upon a number of analyses carried out in settings in which the demand side plays a role (e.g., Nicholson (2002); Dorsey, Jarjoura and Rutecki (2003); Thornton and Esposto (2003); Ku (2011) for the U.S.).[7]

The rest of the paper is organised as follows: Section 2 presents the institutional setting and the data that we use to analyse gender differences in specialty choices. Section 3 motivates our study by providing evidence on the relevance of specialty choice in determining the gender pay gap. Section 4 describes our empirical strategy and main results: it shows that men and women self-select into different medical specialties, and characterises these specialties. Then, section 5 presents our survey data and investigates the existence of gender differences in preferences for job attributes. Section 6 looks into the role of preferences for job attributes in determining gender-based occupational segregation. Finally, section 7 concludes.

2 Institutional Background and Data

2.1 The Medical Studies and Profession in France

The French medical curriculum starts with a highly selective first year, at the end of which all students must sit a national exam that less than 20 percent of the competing students pass, on average.[8] The next two years of the curriculum are dedicated to developing a broad and general set of skills, and the fourth, fifth and sixth years to deepening one's knowledge of the medical sciences and preparing for the National Ranking Examinations (*Épreuves classantes nationales* in French, hereafter NRE), which are held every June and aim to assign each medical student in the country to a residency position (medical specialty, location pair) in which to specialise until the completion of their training.[9]

The NRE is organised by the *Centre national de gestion* (hereafter CNG) an establishment under the supervision of the Ministry of Health and in charge of the recruitment and management of public

[7] Although we tackle different research questions, it is worth mentioning that we are also close to Ketel et al. (2016) in the sense that we rely on the uniqueness of the settings offered by medical schools (in their case, a lottery) to study labour market outcomes later in life.

[8] This strict regulation of entries follows from the large increase in the number of medical students that occurred in the late sixties and early seventies. In 1971, to maintain high levels of earnings in the profession, the government imposed a *numerus clausus* on the number of medical students allowed to continue their studies after the first year. The number of students admitted to the second year reached its lowest level of 3,500 in 1993, and increased steadily until 2020. It was replaced in 2021 by a *numerus apertus* to address a shortage of physicians.

[9] Hereinafter, *subdivision* and *location* are used interchangeably to refer to a geographical area comprising one or more teaching hospitals. The 26 subdivisions of mainland France are shown in Appendix A.1 Figure A1. There are two additional subdivisions in overseas France: Antilles-Guyane and Océan Indien.

hospital staff and practitioners. The centralised exam consists of two and a half days of tests, after which the candidates are ranked solely on the basis of their scores. After the examinations have taken place, the Ministry of Health releases the number of vacancies for each position for that year. Then, on allocation days, the candidates pick a residency position according to their exam rank: the best-ranked candidate chooses first, then so does the second-best, and so on, until all the students have been matched to a position.[10] Candidates observe both the choices of the preceding candidates and the number of vacancies that are available for each position at each point in time. Thus, this application of the Deferred Acceptance algorithm (Gale and Shapley, 1962) ensures that candidates make their decision while being aware of their entire choice set, and therefore elicits the candidates' non-strategic responses.

Upon completion of their sixth year of studies and assignment to a residency position, the newly-appointed residents hold a Master's degree and become an integral part of the medical staff. During their residency, all residents earn the same fixed grid salary and are required to work the number of night shifts and on-call shifts required by their specialty of choice. Specialisation lasts for up to five years, depending on the specialty.

Once they have completed their residency, physicians can work as employees in a public or private medical institution (46 percent in 2017, of which 66 percent work in hospitals), be self-employed (43 percent), or combine employment and self-employment (11 percent). The earnings of employed physicians are usually set by a salary grid based on qualifications and seniority, with the rates set by this grid varying only marginally across specialties. However, self-employed physicians face a more flexible fee-for-service remuneration scheme, their earnings mostly arising from the fees that they charge to patients for each medical consultation or intervention performed. The amount that self-employed physicians are allowed to charge as fees depends on the contractual relationship between them and the national health insurance: regulated and free-billing physicians coexist on the market for the same services. Sector 1 physicians are price takers and must charge the regulated fees set by the national health insurance while sector 2 physicians can charge a markup on top of the regulated fees as long as it is done "with tact and moderation" (Article R4127-53 of the French Public Health Code). The French healthcare system is universal and largely financed by the national health insurance, which is compulsory and contribution-based. Households thus have around 70 percent of their health expenditures covered by public insurance.[11]

[10] Under certain conditions, students are allowed to change specialty during their specialisation. The conditions are (i) to do it only once, (ii) that the change occurs within the same geographical unit, (iii) that the intern could have gotten the newly chosen specialty in that location on their allocation day, given their rank, and (iv) that the change is requested before the end of the fourth semester of the residency, at the latest. From 2010 to 2012, the proportion of students who changed specialty after the NRE allocation ranged from 3.6 to 4 percent of the cohort population only (Golfouse and Pheng, 2015).

[11] Refer to Coudin, Pla and Samson (2015) for a more comprehensive description of the institutional framework.

2.2 Data

The National Ranking Examinations This paper primarily relies on an administrative dataset obtained by combining information published annually by the Ministry of Health via ministerial orders in the Official Journal of the French Republic. It first contains the list of students who sat the National Ranking Examinations each year between 2010 and 2022.[12] It also gives the number of vacancies that are offered each year for each position. Thus, for each exam year, this dataset provides information on the rank and final allocation of each candidate, as well as on the vacancies that are available to them. We are therefore able to identify the occupational choice set faced by each candidate at the time of their choice.

The sample of analysis used in this paper is obtained after imposing the following restrictions. First, we focus on the individuals who are allocated to a residency position.[13] Second, we exclude individuals who choose to work in overseas France, on the grounds that this small group of individuals may differ substantially from the main group. We are left with a sample of $99,076$ individuals across 13 exam years.[14]

Table 1 provides descriptive statistics on the NRE between 2010 and 2022. It shows that on average, $8,188$ vacancies are offered to $8,456$ candidates each year. Close to 60 percent of NRE participants are women and, for the years in which we observe the age of each candidate, they are 25 years old on average. Men taking the exam are slightly older than women, but the population is very homogeneous in terms of age (standard deviation of 2.0).[15] Candidates make their choice when 47 percent of the positions offered that year are still available, on average. The bottom panel of the table focuses on the group of top 5 percent performers, to whom we pay particular attention throughout the paper. It shows that men are more likely to be top performers than women, that top performers are younger than the other candidates, and that top performers face a choice set in which more than 99 percent of all positions are available, on average.

Figure 1 displays the estimated exam score density functions by gender, for the years 2010 to 2022. It shows that men and women differ in their performance in the NRE. While the probability density function for female students has a unimodal distribution, that of male students is clearly bimodal. Men are more concentrated than women both at the top and at the bottom of the exam score distribution, while women are more concentrated between these two extremes.[16] Importantly, both men and women are present in all parts of the exam score support; while the sample is 60 percent female on average, this

[12] Although the current centralised allocation mechanism was first implemented in 2004, we exclude students allocated before 2010 because we do not perfectly observe their final residency position for the years 2004 to 2009. In fact, up until 2010, allocation to specialties was done in two steps and only the first step of the choice was recorded.

[13] Each year, there are fewer allocated individuals than there are exam takers, and this is the case for two reasons. First, once an individual gets their final rank, they may, under certain circumstances, decide not to choose a residency position, and retake the exam the following year. Second, because students take the NRE before knowing if they have met all the requirements necessary to pass the sixth year of medical studies, some of them end up failing it and are therefore not allowed to enter specialisation. All in all, between 88 and 98 percent of takers were allocated to a position during our period.

[14] Appendix A.1 Table A1 shows how the post-selection sample compares to the population.

[15] Note that the 1st, 90th, and 99th percentiles of age are equal to 23, 27, and 34, respectively.

[16] Although this has been the case since 2010, the concentration of men at the top seems to have been slightly more pronounced in more recent years.

Table 1: Descriptive Statistics on the National Ranking Examinations (2010-2022).

	All	Women	Men	Diff.	Obs.
Nb. vacancies	8,189				
	(623.6)				
Nb. takers	8,456	5,000	3,456	1,378	
	(722.8)	(507.1)	(310.8)		
			Full sample		
% Women	0.595				99,076
	(0.491)				
Age	25.2	25.0	25.4	-0.4***	66,146
	(2.0)	(1.8)	(2.2)		
Rank	4,074.9	4,113.7	4,018.0	95.7***	99,076
	(2,449.9)	(2,370.0)	(2,561.8)		
% Available positions	0.470	0.461	0.485	-0.024***	99,076
	(0.312)	(0.304)	(0.323)		
			Top 5 percent		
% Women	0.478				5,082
	(0.500)				
Age	24.2	24.1	24.3	-0.2***	3,395
	(1.1)	(1.0)	(1.2)		
Rank	199.9	205.5	194.8	10.8***	5,082
	(117.2)	(116.3)	(117.8)		
% Available positions	0.992	0.992	0.992	-0.000	5,082
	(0.007)	(0.007)	(0.007)		

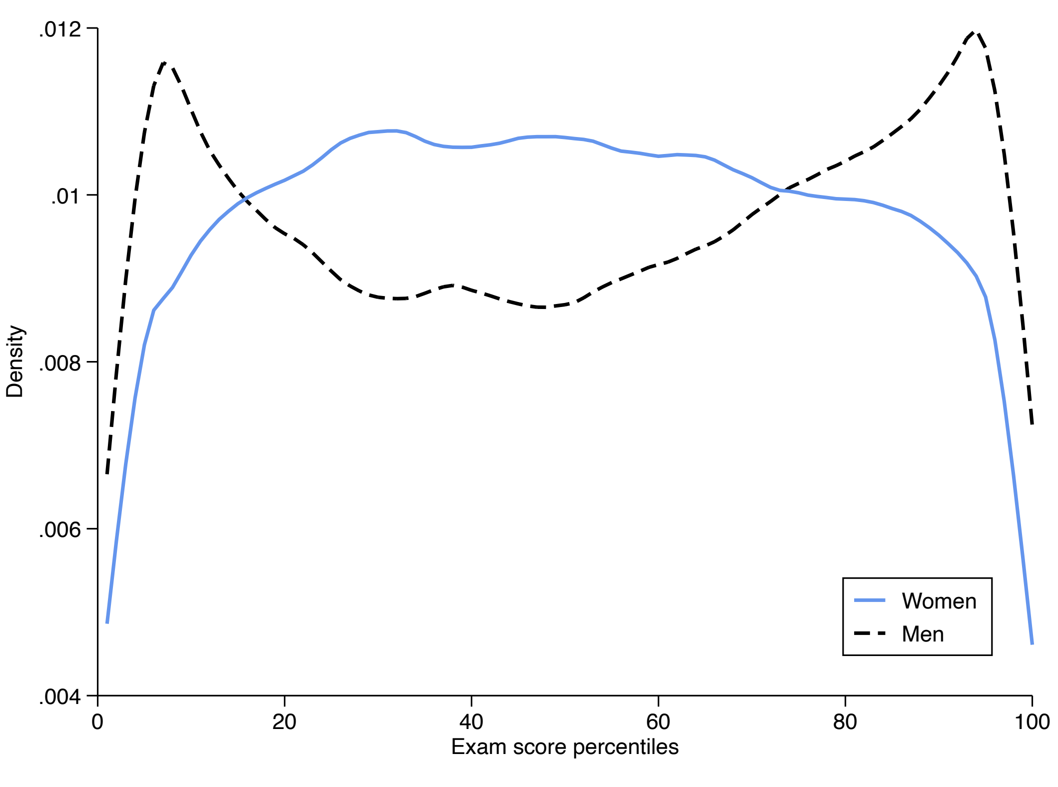
Notes: This table reports descriptive statistics on the National Ranking Examinations between 2010 and 2022, before and after sample selection. *Nb. vacancies* and *Nb. takers* respectively report the number of vacancies that are offered on a given year on average, and the number of candidates taking the exam on a given year on average. "Full sample" refers to the entire population of 2010-2022 NRE takers, while "Top 5 percent" focuses on the top 5 percent exam performers. The table reports the average value for the whole sample (1), the sample of women (2) and the sample of men (3), as well as differences in means between women and men (4) and the number of observations (5). Additionally, standard deviations for the means are reported in parenthesis, and significance levels of the two-sample t-tests for the difference in means between women and men are reported using *** p<0.01, ** p<0.05, * p<0.1.

proportion is lower at both the bottom and at the top of the distribution.

Physicians' Earnings We further need information on the earnings of French physicians, first to motivate the relevance of specialty choice in determining the gender earnings gap in Section 3, and second to characterise the different medical specialties in the second part of Section 4.2. Throughout the paper, we use information on gross yearly earnings and per-procedure fees. In Section 3, we rely on an exhaustive administrative INSEE-CNAMTS-DGFiP file on all physicians with positive earnings from self-employment in 2005, 2008, 2011 and 2014 to compute two measures of earnings: *gross yearly earnings*, which is not adjusted by the quantity of labour, and *per-procedure earnings*, which is.[17] Gross

[17]Data source: *Revenus des libéraux de santé* - 2005, 2008, 2011, 2014, DREES - Ministère de la Santé [producers],

Figure 1: Exam Score Kernel Density by Gender.



Notes: This figure plots the estimated probability density function of the exam score by kernel density estimation for female and male medical students allocated to a residency position between 2010 and 2022.

yearly earnings is defined as the sum of all the fees that are received by a physician for their medical consultations and interventions in a year, while per-procedure earnings refers to the ratio of gross yearly earnings and the number of consultations and interventions performed in that year.[18] The latter aims to proxy hourly wages, which we cannot compute because we do not observe hours worked. Although this impacts the comparability of our pay gap measure with that of other studies, we regard the number of procedures performed as a relevant alternative measure to account for labour supply when analysing the pay gap in the medical profession.

Although this dataset contains very detailed information on self-employed physicians, it does not contain information on physicians who work exclusively under employment contracts and do not bill under self-employment (about two-fifths of all physicians), which we need in Section 4.2 to construct a more comprehensive measure of expected earnings in a given specialty across all types of employment (self-employed and employed physicians). We thus rely on two additional data sources: (i) hospital salary grids from 2016 published by the Ministry of Health in the Official Journal of the French Republic, and (ii) data provided by the Directorate for Research, Studies, Evaluations and Statistics (DREES) on the

ADISP [distributor].

[18]Medical procedures are broadly classified as either consultations or interventions. Within each of these two groups, there exists a much more detailed categorisation of procedures according to a number of criteria (part of the body, action performed, etc.), each of which has its own fee.

demographics of physicians by type of employment and specialty of practice in 2014. Using 2014 data from the INSEE-CNAMTS-DGFiP file, we define expected yearly earnings in a given specialty as the average, cross-sector yearly earnings in that specialty, weighted by the population of physicians in each cell. This measure is not limited to earnings from self-employment: it can be seen as the amount that an intern finishing specialisation and contemplating which type of job to take can earn, in expectation.

Non-Pecuniary Specialty Characteristics In order to further characterise the medical specialties that men and women choose from, we gather additional data on measurable job attributes that matter in the occupational decision. Given data availability constraints, we focus on the following non-monetary attributes.

First, to get a sense of the time requirements of each specialty, we gather data on the average number of hours worked and night shifts performed in each specialty. This information comes from two surveys which were conducted by the largest trade union of medical residents (ISNI) on a representative sample of around 25 percent of the active residents in 2019 and 2012, respectively.[19] We also gather information on the number of years required to complete one's residency program in each specialty from a 2021 ISNI report,[20] and on the proportion of self-employed physicians in each specialty gathered from the 2016 Atlas of Medical Demographics (Conseil national de l'ordre des médecins, 2016), which we use as an additional proxy for time flexibility.

Second, we collect information on perceived job characteristics using the Occupational Information Network (O*NET) database. Following Cortes and Pan (2018), we construct four composite indices capturing the degree of competition, the social orientation, the time pressure, and the interactional skills required by each medical specialty. Appendix A.2 provides extensive details on the measures described in this subsection and on the data sources from which they are derived. It also provides descriptive statistics for each of them.

3 Medical Specialty and the Gender Pay Gap

Despite the decline in occupational segregation over time, there is still a tendency for women and men to choose different types of jobs and specialised training within a given profession (Azmat and Ferrer, 2017). This has been the case for the medical profession in France over the last two decades, where there has been an increase in gender segregation across medical specialties. According to our calculations, the Duncan Segregation Index—capturing the fraction of women (or men) who would have to change specialty to

[19] We acknowledge that the time requirements that are imposed on residents are not necessarily the same as those imposed on physicians within a specialty, but these are the only available databases containing this information at the relevant level. In addition, we argue that they allow us to capture a potentially relevant dimension: the immediate time requirements that individuals face when choosing their specialty.

[20] http://www.futur-interne.com/wp-content/uploads/2021/07/ISNI-GUIDE-2021.pdf.

obtain an identical specialty distribution across genders—among new medical residents was 18 percent in 2004, 19 percent in 2010, and just under 23 percent in 2022. Wasserman (2022) estimates values of the Duncan Segregation Index just above 27 percent in both 2004 and 2010 for a sample of U.S. medical school graduates. Our results highlight a clear trend of increasing specialty segregation by gender among medical students in France and, in turn, in the total population of physicians (from 16 percent in 1999 to 17 percent in 2022).

In this context of increasing gender-based occupational segregation, we start by providing evidence on the importance of specialty choice for the gender pay gap. To do so, we quantify the role that observable factors play in determining the gender earnings gap in the medical profession and show how important specialty of practice is in explaining the gender pay gap compared to other common determinants. We use the administrative individual-level data file described in Section 2.2, which covers more than half of all active physicians in these years.

We focus on physicians working in the regulated-fee sector, known as sector 1, because their earnings can be tightly linked to their productivity through the number and type of procedures they perform. Importantly, these physicians' earnings are not tied to tenure (this sector does not feature a promotion scheme), there are no convex returns to hours worked (physicians are paid at the same rate whether they perform 20 or 40 consultations a day), or pay differentials based on negotiations. Moreover, using physicians practising in sector 1 also allows us to abstract from sorting across employment types and sectors.[21] However, acknowledging the gender-based sorting into the different types of employment which exists in the medical profession,[22] we also perform the analysis on sector 2 physicians and report the results in Appendix B Table B2.

Table 2 reports the female coefficient from estimating a set of standard Mincer regressions of the form:

$$ln(Earnings_{it}) = \beta_0 + \beta_1 Female_i + X'_{it}\rho + \epsilon_{it} \quad (1)$$

for physician i in year t, where $Earnings_{it}$ is per-procedure earnings or gross yearly earnings, $Female_i$ indicates that physician i is a woman, and X_{it} represents the set of control variables described below.

Table 2 reports the results on per-procedure earnings for the sake of exposition, and Appendix B Table B1 contains further results on per-procedure earnings and yearly gross earnings. Column (1) shows that after controlling for year and size of the urban area of practice (UA), female physicians established in the regulated-fee sector (sector 1) earn about 9 percent less for each procedure they perform than their male

[21] Once a physician has established their practice in sector 1, they cannot switch to sector 2. However, the reverse switch is possible.

[22] For instance, according to the DREES Open Data available at https://data.drees.solidarites-sante.gouv.fr/, 52 percent of female physicians worked under employment contracts in 2014, compared to 32 percent of men. Similarly, female self-employed physicians are more likely to be established in the regulated-fee sector 1 than in the free-billing sector 2 (76 percent work in sector 1) compared to men (74 percent), as shown in Appendix A.2 Table A2.

counterparts.[23] This indicates that most of the 34 percent gap in yearly earnings shown in column (1) of Panel B of Appendix B Table B1 can be explained by the fact that women perform fewer procedures. The pay gap is even wider in the free-billing sector (sector 2), where women earn on average 24 percent less than men for each procedure performed, and 38 percent less than men for their yearly earnings. We then incrementally add individual- and practice-level control variables and report the results in columns (2) to (4). These controls only slightly reduce the estimated gender gap: after controlling for age, experience, practice type and procedure type, female physicians in the regulated-fee billing sector still earn 7 percent less per procedure than their male counterparts.[24]

Table 2: Female-to-Male Earnings Gap in the Regulated-Billing Sector.

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.088***	-0.098***	-0.084***	-0.069***	-0.039***	-0.038***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Mean dep. variable	3.484	3.484	3.484	3.484	3.484	3.484
Adj. R-squared	0.055	0.101	0.150	0.291	0.468	0.492
Observations	312,153	312,153	312,153	312,153	312,153	312,153
Year & UA size FEs	✓	✓	✓	✓	✓	✓
Age & experience	-	✓	✓	✓	✓	✓
Type of practice	-	-	✓	✓	✓	✓
Procedure type	-	-	-	✓	-	✓
Specialty FEs	-	-	-	-	✓	✓

Notes: OLS estimates of the gender gap in the per procedure earnings estimated from equation (1) using all physicians with positive gross earnings billed in Sector 1. Per-procedure earnings refers to the ratio of gross yearly earnings (sum of all the fees received for medical consultations and interventions) and the number of consultations and interventions performed in that year. "Year & UA size FEs" are fixed effects for the size of the urban area in which the practice is located, "Age & experience" includes indicators for 5 age categories as well as the number of years of experience as a private practitioner and its squared value, "Type of practice" refers to a set of dummy variables for the legal status of the practice and the share of yearly earnings arising from self-employment, "Procedure type" is the share of yearly procedures that are consultations, and "Specialty FEs" refers to 16 indicators variables for each medical specialty. Heteroskedasticity-robust standard errors reported in parenthesis, ***, **, * denoting significance at the 1%, 5% and 10% levels, respectively.

Columns (5) adds medical specialty fixed effects, which reduce the remaining gender pay gap per procedure to 3.9 percent, implying that 43 percent of the residual gender pay gap per procedure can be explained by women and men sorting into different specialties. This makes specialty of practice the most important observable factor explaining the gender pay gap.[25] In the free-billing sector, the reduction is even larger, with specialty of practice explaining 61 percent of the residual gender pay gap, leaving the gap at 9 percent.

[23]The corresponding unconditional gender pay gaps in per-procedure earnings and gross yearly earnings are reported in column (1) of Appendix B Table B1. They are respectively equal to 7 percent (Panel A) and 34 percent (Panel B).

[24]In order to account for different fees and fee compositions in each specialty, in column (4), we control for the share of procedures performed that are considered to be consultations as opposed to interventions. Since different specialties require different types of procedures, this control may already partially account for gender differences in specialty sorting.

[25]Although specialty of practice is also the most important observable determinant of the gender gap in yearly earnings (Panel B of Table B1), controlling for it does not close the gap. This is mainly due to the fact that gender differences in labour supply are not taken into account when using gross yearly earnings instead of per-procedure earnings.

The gender pay gap we uncover in the regulated-fee sector is two-thirds of the gender gap in hourly earnings found in a sample of young U.S. doctors in 1990 (Sasser, 2005), while the gender gap in the free-billing sector is almost twice as large. We did expect the gender pay gap to be smaller in our setting than in Sasser (2005) because of the billing regulations imposed on sector 1 physicians, it is however surprising that the gender pay gap we uncover among physicians in the free-billing sector is almost twice as large as that in the U.S. Our findings are very similar to those of Blau and Kahn (2017) for a sample of U.S. workers over several decades, in which the authors report "a continued substantial role for occupation and industry in explaining the gender wage gap". Our results also mirror those found for other homogeneous groups of workers such as recent MBAs (Bertrand, Goldin and Katz, 2010), pharmacists (Goldin and Katz, 2016), lawyers (Azmat and Ferrer, 2017), and Uber drivers (Cook et al., 2021). Specifically, the latter study documents a remarkably similar gender earnings gap of around 7 percent in this "gig" economy, where pay is set in a similar way as in the regulated-fee sector for physicians.

We interpret these results as clear evidence of the importance of specialty choice for the labour market outcomes of medical students. In the remaining sections, we seek to understand the reasons behind the gender differences in specialty choices and how these are partly driven by supply-side factors.

4 Gender Differences in Specialty Choices

In this section, we start by describing the empirical strategy that we use to analyse gender differences in specialty choices. It relies on two main features. First, given that in this framework individuals sequentially choose their future medical specialty from a finite list of positions and that employers have no decision-making power over the match, after conditioning on individual choice sets for residency positions, gender differences in choices are entirely attributable to supply-side factors. Second, focusing on candidates at the top of the exam performance allows us to make sure that choices are unconstrained by position availability, which further ensures that specialty sorting is largely driven by preferences. This is explained in details in Section 4.1. We then show our estimation results in Section 4.2.

4.1 Empirical Strategy

To isolate the role of jobseekers' preferences in determining gender-based occupational segregation, the ideal experiment would compare the occupational choices of individuals with similar characteristics (education, experience, ability, etc.), who face no screening by employers and the same occupational choice set, and who differ only in their gender. The setting provided by the National Ranking Examinations is very close to this ideal. Not only does it remove the role of employers in the hiring process, but it also allows us to observe, for each year, (i) each individual's specialty choice, (ii) the rank of that individual in the population of candidates, and (iii) the number of vacancies offered per position. Taken together, these elements allow us to retrieve each candidate's exact choice set, and thus to compare the decisions of candidates facing the same choice sets.

To further identify the role of jobseekers' preferences, we pay particular attention to the group of top performers in the national exam, for whom most preferred and revealed residency choices coincide. Given the sequentiality of choices imposed by the NRE ranking, these top performers are the first to choose their residency position during the national allocation process. At the time of their choice, *all* positions are still available. As a result, these students are entirely unconstrained by position availability, which ensures that their revealed preferences for specialty coincide with their most preferred specialty.[26] However, only 554 individuals are fully unconstrained over the entire period. Therefore, we focus instead on the top 5 percent NRE performers (amounting to $5,082$ individuals). At their time of choice, these individuals face a choice set composed of 99.2 percent of all the offered positions, on average (Table 1). Importantly, all the specialties and subdivisions have at least one remaining vacancy throughout the top ventile of the performance distribution,[27] which is evidence of the fact that virtually all of these top performers are unconstrained by position availability when choosing their residency position.[28] Appendix C.2.2 shows that our results are robust to a number of alternative definitions for being unconstrained by position availability.

We estimate the following equation by ordinary least squares at different points of the exam rank distribution:

$$y_i = \beta female_i + \gamma_{c(i)} + \epsilon_i \tag{2}$$

where i refers to a candidate in a given year, y_i is an indicator variable taking the value 1 if candidate i chooses the specialty under consideration and 0 otherwise, $female_i$ is an indicator variable taking the value 1 if candidate i is a woman and 0 otherwise, and $\gamma_{c(i)}$ are choice set fixed effects.[29] We define choice set fixed effects as groups of five individuals with consecutive ranks.[30] They allow us to control

[26]We further argue that this group of individuals is of particular interest as one might expect gender differences in labour market aspirations to be limited in this group of highly-educated highly-motivated individuals.

[27]In some years, some positions are filled as the top 5 percent candidates are still choosing. For instance, in 2022, 30 positions were filled before the 434 top 5 percent candidates had made their residency choice. The first position to be filled was Hematology in Bordeaux, as it only offered one vacancy that year. As a matter of fact, 50 percent of the positions that got filled within the top 5 percent were in Hematology, Maxillofacial Surgery, Plastic Surgery, or Nephrology. To reassure the reader of the fact that, despite these filled positions, the top 5 percent performers are virtually unconstrained by position availability, we compare the stated and revealed choices of our 2022 survey respondents and check whether discrepancies are due to positions being filled already. We find that out of the 75 top performers among our survey respondents, only one had declared a preferred position that was filled by their time of choice.

[28]We acknowledge that exam performance is likely to result from both ability and effort, the latter being driven by specialty preferences. However, what matters for our analysis is that the ranking perfectly defines choice sets, regardless of how that ranking was formed.

[29]We prefer OLS over a Probit or Logit regression framework to avoid bias arising from the well-documented incidental parameter problem (Neyman and Scott, 1948). Furthermore, since the job attributes described in Section 2.2 vary at the speciality level rather than the individual level, their parameters would not be identified in a multinomial logistic regression framework.

[30]Appendix subsection C.2.1 shows that our results are robust to different definitions of choice set fixed effects.

for all the characteristics that are common to individuals with virtually equivalent positions in the rank distribution, such as exam performance (which can be seen as a combination of individual ability and effort) and, most importantly, the set of available positions to choose from.

Appendix A.1 Figure A2 provides an overview of the vacancy filling process for each specialty over the exam score distribution. It shows how some specialties are filled earlier than others (because of their popularity and number of available vacancies), and justifies conditioning on choice set to estimate gender differences as well as focusing on candidates at the top of the exam score distribution. The results obtained on the group of top 5 percent performers are displayed in Figure 2, while the estimates obtained along the entire support of the performance distribution are presented in Appendix C.1.

4.2 Results

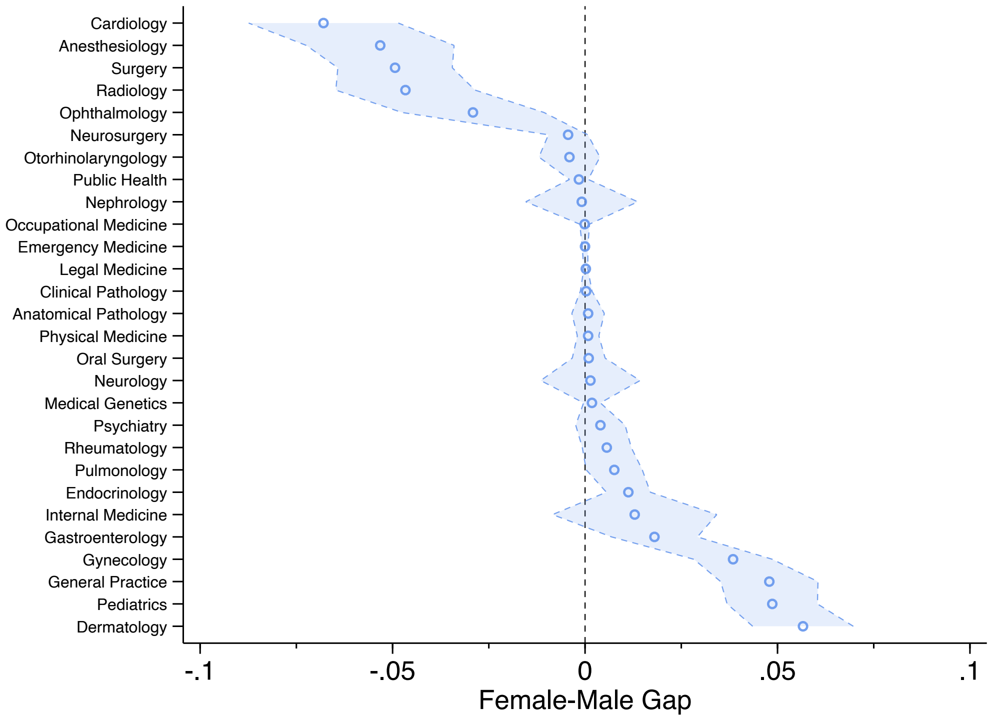
Self-Selection into Specialties Figure 2 plots gender differences in the propensity to self-select into each specialty estimated on the top 5 percent exam performers. Precisely, each row reports the coefficient $\hat{\beta}$ and its 95% confidence interval estimated from equation (2) using an indicator variable taking the value 1 if candidate i chooses the specialty under consideration and 0 otherwise. As a result, each point estimate can be interpreted as the gender gap in the probability of self-selecting into a given specialty among students who are unconstrained by position availability.

Figure 2 presents the main result of the paper: men and women facing the same occupational choice set make very different choices. Women at the top of the performance distribution are 6.8 percentage points less likely to choose cardiology, 5.3 percentage points less likely to choose anesthesiology, and 4.9 percentage points less likely to choose surgery than their male counterparts facing the same choice set. Symmetrically, they are 4.8 percentage points more likely to self-select into general practice, 4.9 percentage points more likely to choose pediatrics, and 5.7 percentage points more likely to choose dermatology than their male counterparts. These results are striking, given that top-performing individuals can choose any specialty in any location, and thus do not face the specialty-location trade-off that some candidates with lower ranks face. Our results imply that men and women facing the same choice set *prefer* different medical specialties: even when they are unconstrained by position availability and do not have to go through a hiring process, men and women still choose to work in different specialties. It indicates that differences in preferences for job characteristics are a key component of the observed gender differences in self-selection into medical specialties.[31]

Figure 2 further shows that there are specialties for which no gender differences are observed: they

[31] Ku (2011) shows in the U.S. context that women are more likely than men to choose general practice, pediatrics, and gynecology, while men are more likely to choose surgical specialties and the E-ROAD specialties (emergency medicine, radiology, ophthalmology, anaesthesiology, and dermatology). By focusing on a setting in which only supply-side factors are at play, we go one step further and show that some gender differences in specialty choice are still present when focusing on individuals who are unconstrained by position availability and search and matching frictions. Our results also suggest that pooling the E-ROAD specialties together hides some gender differences in self-selection.

Figure 2: Gender Differences in Specialty Choices (Top 5% Exam Performers).



Notes: This figure plots the coefficients $\hat{\beta}$ resulting from estimating (2) for each medical specialty, using the sample of students who scored in the top 5% of the National Ranking Examinations in the years 2010-2022 (5,078 individuals). This group of students chose their residency position when 99.7% or more of all positions were still available. Negative (positive) values indicate that, on average, women are less (more) likely to self-select into the corresponding specialty than men with a virtually identical choice set. All the gender gaps are estimated using OLS and include choice set fixed effects defined as groups of 5 individuals with consecutive exam scores. The shaded area show 95% confidence intervals computed using heteroskedastic robust standard errors.

exhibit gender gaps that are very close to 0, albeit most of them are not statistically significant at the 5 percent level. This is because, under our regression framework, the magnitude of each specialty's estimated gender gap is bounded by its relative popularity.[32] To understand how the estimated gaps change when accounting for each specialty's relative (female) popularity, Appendix C.1 Figure C4 plots the point estimates from equation (2) expressed as the proportion of the share of women choosing each specialty. The relative popularity of each specialty in the population of top 5 percent exam performers is also shown by its marker's size. Although the magnitude and interpretation of the gender gaps change with this transformation, the order of the specialties and the conclusions remain unchanged.[33]

Gendered Self-Selection and Job Characteristics We next correlate the gender gaps estimated on the population of top 5 percent performers with the different job characteristics defined in Section 2.2 and show the results in Figure 3. In each sub-figure of Figure 3, a circle represents one of the 28 medical specialties under study and its size shows the specialty's relative popularity in the population of top five

[32]Taking a fictitious specialty that is only selected by 1 percent of the students, this means that the estimated gender gap cannot be larger than 1 percent.

[33]In fact, the specialties that see their gender gap increase the most (in absolute terms) are among the least popular, and their gender gaps remain statistically insignificant. The one exception is endocrinology, which exhibits a positive and significant gender gap in Figure 2 as well.

percent exam performers. The dashed line plots the prediction from a linear regression of the gender gap on the characteristic. All job characteristics have been normalised to have a mean of zero and a standard deviation of one in the sample of medical specialties. Recall that a positive female-male gap estimated from equation (2) means that women are more likely than men to self-select into a given specialty, while a negative gap means that men are more likely than women to choose the specialty under study. The figure shows clear correlations between self-selection patterns and medical specialty characteristics. Male students at the top of the performance distribution are more likely than their female counterparts to self-select into medical specialties that offer higher gross yearly earnings (unadjusted for labour supply) and higher per-procedure earnings. They are also more likely to select specialties that are more competitive and more demanding in terms of time requirements (more night shifts, more time pressure, longer residency programs, and to a lesser extent more hours worked).

So far, we have shown that very similar men and women facing the same choice set choose different occupations, and that these occupations differ in terms of earnings and non-pecuniary amenities. It suggests that preferences for job characteristics play an important role in explaining gender-based occupational segregation. This is what we document in the remaining sections using our survey data, first by looking at gender differences in preferences for job attributes, and then by providing evidence on their role in determining occupational sorting.

5 Preferences for Job Characteristics

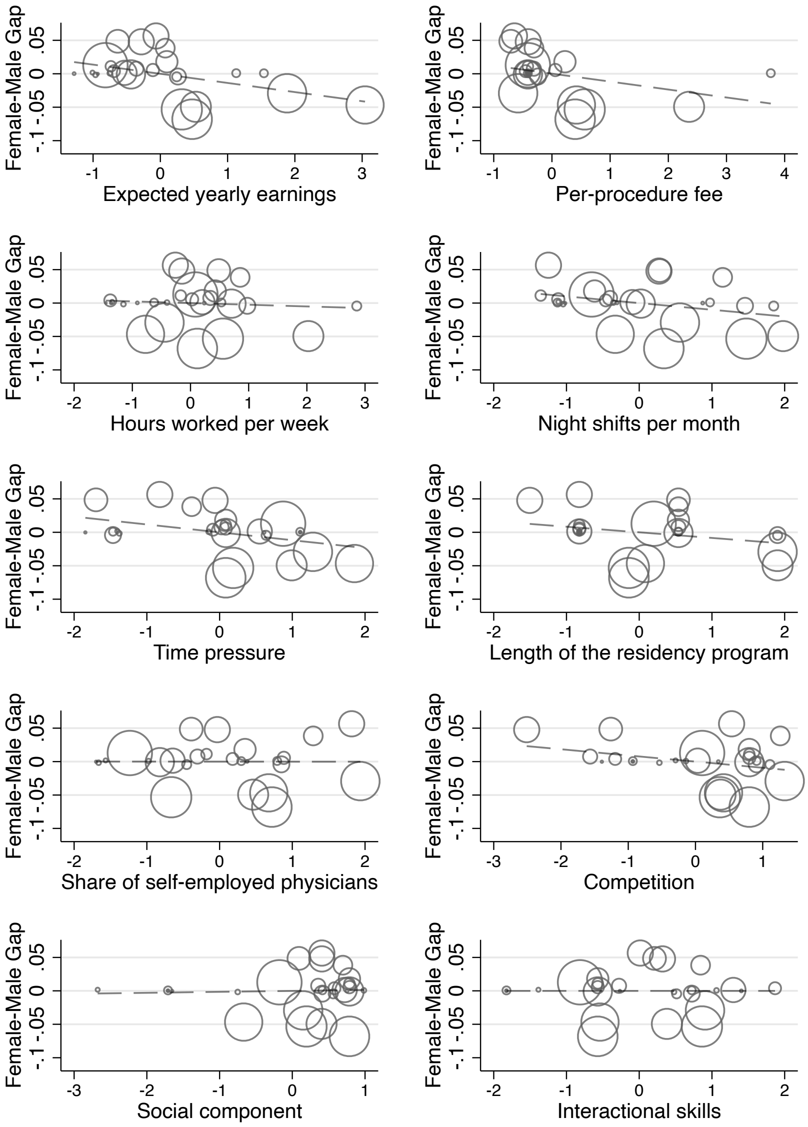
In this section, we present our survey of a sample of candidates for the National Ranking Examinations, in which we elicit preferences for a range of job attributes that could explain the gender segregation in occupations documented above. We then examine the existence of gender differences in preferences for job characteristics.

5.1 Survey to Medical Students

To obtain evidence on the drivers of the observed gender differences in occupational choices, we conducted two online surveys of a sample of prospective NRE candidates. Importantly, for those respondents who consented, we also obtained their rank in the national exam and their actual choice of specialty and location. The recruitment of medical students was carried out under the MEDSPE research project (N°ANR-21-CE26-0013-01). 8,043 of the 9,070 students who registered for the 2022 National Ranking Practice Examinations (*Épreuves Classantes Nationales préparatoires*, hereafter NREp, a mock exam aimed at training for the NRE) were invited to participate in a first survey.[34] This corresponds to 83.7

[34]The NREp were held between 21 and 23 March 2022 on a national digital evaluation platform known as SIDES, which stands for *Système informatisé distribué d'évaluation en santé*. In addition to hosting the national practice (or mock) exam, it also provides material to prepare for the NRE. The difference between the number of NREp participants and the number

Figure 3: Estimated Gender Gap and Job Characteristics.



Notes: Each graph shown in this figure plots the gender gaps estimated from equation (2) on the sample of top 5 percent exam performers against a given job characteristic as well as the prediction from a linear regression of the gender gap on the characteristic. Each graph thus shows the correlation that exists in our sample between the self-selection gender gap and the job attribute under study. Each marker represents one of the 28 medical specialties of the sample. The size of each marker represents the relative popularity of each specialty in the population of top 5 percent exam performers. The job characteristics have been normalised to have a mean of zero and a standard deviation of one in the sample of medical specialties.

percent of the population of 2022 sixth year medical students. This first survey, developed with oTree (Chen, Schonger and Wickens, 2016), was conducted between 8 and 20 March 2022, just before the NREp, and gathered 3,525 respondents. Of these, 2,588 (73 percent) agreed to be contacted again for follow-up surveys. Of these 2,588 individuals, 1,382 participated in the follow-up survey. The follow-up survey data were collected using Qualtrics, a web-based survey tool, over a four-week period in May 2022. The participation period began 45 days and ended 15 days before the 2022 NRE.

of students that were contacted to take part in the first survey is due to the fact that the students' contact details were collected on a date when not all NREp participants had registered yet.

The surveys took approximately 28 minutes and 25 minutes to complete, respectively. In the first survey, respondents were incentivised with a fixed payout of 10 euros and a variable payout based on performance in different tasks, as well as a raffle with several cash prizes. In the follow-up survey, respondents were offered the chance to enter a lottery to win one of 17 prizes with a total value of 1,462 euros. A timeline showing the different stages of this project (NRE dates, survey dates) is presented in Appendix D Figure D1. In this paper, we rely almost exclusively on data collected during the follow-up survey.

In addition to questions on demographics and family background, the follow-up survey consisted of several sections. First, respondents were asked about their preferences in terms of residency positions and their expected performance in the NRE. They were also asked about their knowledge and expectations of the working conditions (average earnings, number of hours worked including night shifts, proportion of women, and proportion of self-employed physicians) in their preferred specialty and in two other randomly-selected specialities, one of which is predominantly selected by women and the other by men (defined base on Figure 2). The main section of the follow-up survey consisted in a hypothetical job choice experiment, as described in Section 5.2. We complemented the information collected through the hypothetical job choices by asking respondents how important a series of characteristics were in their choice of residency position (see Section 5.3 for a detailed analysis of these survey questions). Finally, we asked respondents about their experiences and expectations of gender-based discrimination and elicited their beliefs about gender norms. The information on respondents participation and rank in the national exam, as well as their actual residency choice was merged with their survey responses.

Table 3 summarises the information available jointly for the students who were contacted and responded to the surveys and for those who participated in the NREp and the NRE. Both surveys received an acceptable response rate, 43.8 and 53.4 percent respectively, with female students being slightly more likely to participate in both surveys. More than 98 percent of our respondents participated in the 2022 National Ranking Examinations, and those with higher scores were more likely to participate in both surveys. For instance, 29.3 percent of the respondents ended-up in the top quartile of the NRE rank distribution, while only 18.6 percent did so in the bottom quartile.

Appendix D Table D1 provides some descriptive statistics on the men and women of our sample. Women are 24.3 years old and 4 months younger than men on average. 99 percent of the women of the sample are French, compared to 95 percent of the men. While 39 percent of our men and women are single at the time of the follow-up survey, women are slightly more likely than men to be married or in a civil partnership, but the difference is not statistically significant. There are also significant differences between the fertility plans of our male and female respondents: while 30 percent of women declare wanting to have children in the five years following the survey, only 22 percent of men do.

Table 3: Students' Characteristics Across Samples.

	1st survey			2nd survey		
	Contacted	Respondents	National Mock (NREp)	Contacted	Respondents	National Exam (NRE)
Women	5,085 (63.2)	2,374 (67.3)	5,722 (63.1)	1,737 (67.1)	934 (67.6)	5,990 (64.5)
Men	2,958 (36.8)	1,151 (32.7)	3,348 (36.9)	851 (32.9)	448 (32.4)	3,292 (35.5)
National Exam Rank						
Take-up rate	97.2	98.2	96.9	98.3	98.5	100.0
1st quartile	26.1	27.2	25.2	27.7	29.3	25.0
2nd quartile	22.6	21.9	23.3	21.4	20.0	25.0
3rd quartile	21.3	19.8	22.5	19.5	17.4	25.0
4th quartile	23.6	20.3	25.1	20.2	18.6	25.0
Contact and response rates	88.7%	43.8%		73.4%	53.4%	
Total	8,043	3,525	9,070	2,588	1,382	9,282

Notes: This table shows the total number of students in each sample (1st survey, NREp, 2nd survey, NRE), as well as their repartition across genders and NRE rank quartiles. The different samples are displayed in chronological order. The NRE sample contains the entire population of NRE candidates. Contact rates are computed by dividing the total number of students who received an invitation to participate in either survey by the total number of students in the corresponding sample.

Turning to the respondents' educational background, virtually all of them followed the scientific path in high school, and there is a higher proportion of women who passed the high school final exam with the highest honors. Finally, in our sample, men are as likely as women to pass the first year of medical school at the first attempt. Similarly, there are no significant differences in the propensity of men and women to obtain a top or bottom rank in the medical school first year final exam.[35] Additionally, Appendix D Table D2 reports the geographical distribution of respondents by gender based on the location of their university, and Figure D2 compares their self-reported expected exam performance with their actual performance.

5.2 Valuations of Job Attributes: A Hypothetical Job Choice Model

We use a hypothetical job choice methodology to estimate individual valuations of different workplace attributes that characterise healthcare occupations. The setting under study is particularly well suited to the collection and analysis of hypothetical job choice data. First, we survey students only a few months before they start their first job (their residency program), which minimises the influence of previous work experience on preferences.[36] This ensures that preferences for job attributes should not change drastically between their elicitation and the individual's actual labour market decision. Second, we link each student's estimated preferences for workplace attributes to their actual specialisation choice. Note that we do this in a setting that is incentive-compatible and not plagued by common labour market frictions. In doing so, we improve upon previous studies in which the respondents' labour market decisions typically occur several years after preference elicitation, and in which labour market frictions may prevent individuals

[35] Note that female and male respondents also differ in terms of parental education: women are more likely than men to have parents with some level of tertiary education.

[36] All medical students are required to complete short rotations in core medical specialties as part of the medical curriculum. Therefore, our elicited preferences for job attributes reflect each individual's personal experience of these placements.

from matching with their preferred job (e.g. Wiswall and Zafar, 2018).

5.2.1 Design of the Experiment

Survey participants were presented with an introductory text with the following instructions:

> Imagine that **you have just finished your training** in your preferred residency position and that you now need to choose between different **salaried jobs** that are offered to you.
>
> Each job is characterised by [a list of 4 job characteristics].
>
> **These jobs are identical** in all respects, including the medical specialty that is practiced, except for the four characteristics mentioned above.
>
> These jobs are presented to you in the 8 scenarios below. In each scenario, we ask you to indicate **the probability that you pick each of the jobs**. The probability attributed to each job must be a number from 0 to 100, and the sum of probabilities within a scenario must be equal to 100.

Following similar designs used in other papers (Blass, Lach and Manski, 2010; Eriksson and Kristensen, 2014; Mas and Pallais, 2017; Maestas et al., 2018), and in particular the one used in Wiswall and Zafar (2018), we presented respondents with two sets of 8 scenarios (16 job scenarios in total). Each scenario consisted of three hypothetical jobs that were constructed by exogenously varying four job characteristics. Respondents were asked to provide choice probabilities for each of the three jobs, rather than a discrete answer. This allowed them to express uncertainty about their choice and to rank the alternatives in order of preference.[37]

In the first set of 8 scenarios, we varied (i) monthly earnings, (ii) number of hours worked, (iii) number of night shifts and (iv) having a predictable fixed schedule. In the second set of scenarios, we varied (i) monthly earnings, (ii) earnings growth after 10 years, (iii) commuting time, and (iv) proportion of women in the workplace. Following Wiswall and Zafar (2018), we decided not to vary these 8 job characteristics at once, so as not to overwhelm respondents. We randomised the order in which the two sets of scenarios appeared. Importantly, respondents were told that the jobs were identical in all respects except for the four characteristics that were shown to them. The text also emphasised that all hypothetical jobs were referring to the same medical specialty.

We selected the workplace attributes described above based on prior literature, on prior evidence related to the healthcare sector, and because they showed considerable variation both across specialties

[37] As explained in detail by Blass, Lach and Manski (2010), this "elicited-choice-probabilities" approach to estimating individual preferences for job attributes has the drawback that it requires respondents to have some knowledge of probabilistic reasoning and requires more effort than the "stated-choice" approach. Given that medical students are familiar with probabilities and statistics since it is part of their medical curriculum, the former should not be a major concern. To alleviate it further, we asked respondents to answer a test scenario before moving on to the actual task. To tackle the second concern, we identify and exclude individuals who completed the experiment in three minutes or less and thus appear not to have paid attention to the task, as explained in the following section.

and across jobs within a speciality. Earnings and earnings growth were included because they have been shown to be important determinants of career choice (Thornton and Esposto, 2003; Gagné and Léger, 2005; Wiswall and Zafar, 2018).[38] Earnings growth after 10 years additionally captures the respondents' intertemporal preferences and their valuation of career progression. Number of hours worked, number of night shifts, and schedule predictability were included because they relate to work-life balance and to the job's remuneration structure and to the gender earnings gap (Goldin, 2014; Cortes and Pan, 2018).[39] Having a predictable schedule captures the ability to know what time of day one is expected to work, which should correlate with planning ahead and having a better work-life balance (Thornton and Esposto, 2003; Harris et al., 2017; Mas and Pallais, 2017). Commuting time was included because it has been shown to be a relevant factor in one's occupational decision, especially for women (Harris et al., 2017; Le Barbanchon, Rathelot and Roulet, 2020). Finally, we included the proportion of women in the workplace to assess whether female composition matters differently for men and women.

5.2.2 Model and Estimation

Under the assumptions of linear preferences over the job attributes and of utility-maximising decision makers, we use the canonical random-utility model in which medical student i derives utility from choosing job j in the following way:

$$U_{ij} = \beta_i X_j + \epsilon_{ij} \tag{3}$$

where X_j is a vector containing the observed attributes of job j, β_i is a vector containing individual i's utility weights over those attributes, and ϵ_{ij} is an idiosyncratic error term containing the remaining characteristics that affect respondent i's utility (it is observed by the respondent but not by the researcher).

We estimate the following linear median regression model:

$$M\left[\ln\left(\frac{p_{ij}}{p_{i1}}\right) \mid X\right] = (X_j - X_1)'\beta_i \tag{4}$$

where p_{ij} is the probability that individual i chooses job j, and $M[\cdot]$ is the median operator. The assumptions required to estimate the utility weights without measurement error and the steps leading to equation (4) are presented in detail in Appendix E.1.

After replacing 0 probabilities by 0.001 and 100 probabilities by 99.9 to avoid log odd ratios that

[38] We slightly depart from Wiswall and Zafar (2018) by presenting respondents with monthly earnings rather than annual earnings. We do this because informal discussions with workers in the French healthcare sector revealed that this is the way workers think about earnings.

[39] Respondents were instructed that the night shifts measure also included on-call time. They were also instructed that the measure of hours worked *did not* account for the number of night shifts shown in the same job.

equal minus and plus infinity, we estimate equation (4) by Least Absolute Deviations (LAD) for different specifications of the attributes.[40] Our preferred specification linearly includes all the job characteristics described above.[41] To avoid making assumptions about the population distribution of preferences, we estimate β_i separately for each respondent i using 32 unique observations. Based on these estimates, we then obtain population statistics and use cluster bootstrap samples with replacement to conduct inference.

We also compute individual-level willingness-to-pay estimates, which are easier to interpret than the β_i preference parameters estimated from equation (4). We denote individual i's willingness to pay for a change in the level of attribute X_k from $X_k = x_k$ to $X_k = x_k + \Delta$ as $\text{WTP}_{ik}(\Delta)$. Given the assumed linear utility function, $\text{WTP}_{ik}(\Delta)$ is defined such that:

$$x_k \beta_{ik} + \beta_{i1} \ln(Y) = (x_k + \Delta)\beta_{ik} + \beta_{i1} \ln(Y + \text{WTP}_{ik}(\Delta)) \tag{5}$$

where β_{ik} is individual i's preference for attribute k, β_{i1} is their preference for earnings, Δ is the unit increase in attribute k, and Y is the level of earnings. As a result:

$$\text{WTP}_{ik}(\Delta) = \left[\exp\left(\frac{-\beta_{ik}}{\beta_{i1}}\Delta\right) - 1\right] \times Y \tag{6}$$

Thus, $\text{WTP}_{ik}(\Delta) > 0$ indicates that individual i is willing to pay to avoid increasing the "bad" attribute k by Δ, while $\text{WTP}_{ik}(\Delta) < 0$ means that individual i is willing to pay to get Δ units more of the "good" attribute k.

Of the $1,382$ survey respondents, 90 percent completed the 16 scenarios of the hypothetical job choice experiment ($1,244$ individuals). Appendix E.2 describes the choice probabilities elicited during the exercise before sample selection. To focus our analysis on individuals who were attentive during the exercise, we implement the following sample restrictions. We first remove responses from individuals who spent less than 3 minutes on the job choice experiment (20 individuals).[42] Second, we exclude the 27 individuals who have lexicographic preferences for a certain job attribute, for whom, from a theoretical point of view, there is no trade-off between job characteristics. Finally, individuals with at least one estimated willingness to pay in the bottom or top 1 percent are also excluded from our analysis. After implementing these restrictions, we are left with the elicited probabilities of $1,146$ medical students.[43]

[40] The LAD model is invariant to transformations of the extreme values (0 and 100) that do not alter the ordering of preference values around the median.

[41] We report results from alternative specifications in Appendix E.3 Figures E4 to E6.

[42] Appendix subsection E.2 shows that these individuals are much more likely to input extreme probabilities than the rest of the respondents, which we take as evidence of their inattention.

[43] Results are robust to using the whole sample of respondents who completed the job choice experiment. We nonetheless decide to restrict the sample in an attempt to get as close as possible to a convincing, trustworthy sample. Note that this restriction also implies clearer graphical exposition.

5.2.3 Results

We estimate kernel densities of the job preference parameters β from job choice model (4) for each gender and report them in Appendix E.3 Figure E3, together with some descriptive statistics on the estimates in Table E1. Unsurprisingly, most respondents prefer jobs with higher earnings, and a significant share of respondents have a strong taste for higher earnings (bunching at around 25 log points). We also find that a non-negligible number of respondents have zero or negative preference estimates for earnings.[44] The mean and median values of the other preference parameters have the expected signs.

The distributions of the willingness-to-pay estimates for each gender expressed as a percentage of average earnings across all scenarios, namely $3,511$ euros per month, are displayed in Figure 4.[45] We complement this figure with some descriptive statistics on the estimates in Appendix E.3 Table E2.[46]

The estimates displayed in Figure 4 show that the men and women of our sample have remarkably similar willingness to pay for the job attributes under study. The two-sample Kolmogorov–Smirnov tests reported on each sub-figure (KS) show that we fail to reject the null hypothesis of equality of the distributions for all the job attributes, except for commuting time which appears to be marginally significant. However, a closer look at the other p-values reported on the sub-figures sheds light on some heterogeneity: the distribution of estimated WTPs for hours worked is more skewed to the right for women than for men (1.8 vs. 1.5 percent of monthly earnings required to work for an extra hour at the median, and 3.6 vs. 2.9 percent at the 75^{th} percentile). This suggests the existence of a group of women who require very large compensations to work one extra hour per week. More generally, gender differences seem to be more pronounced in the tails of the WTP distributions. It is important to note that our estimates capture preferences at a particular point in the life cycle of the survey respondents and that preferences for job attributes may evolve over time. Nevertheless, we argue that the timing of our survey allows us to measure preferences at a very relevant moment for medical students, just before they choose a medical specialty that will determine their field of work for the rest of their career. Next, we compare these results to those obtained with a simpler methodology to retrieve preferences for job characteristics.

5.3 Valuations of Job Attributes: Self-Reported Measures

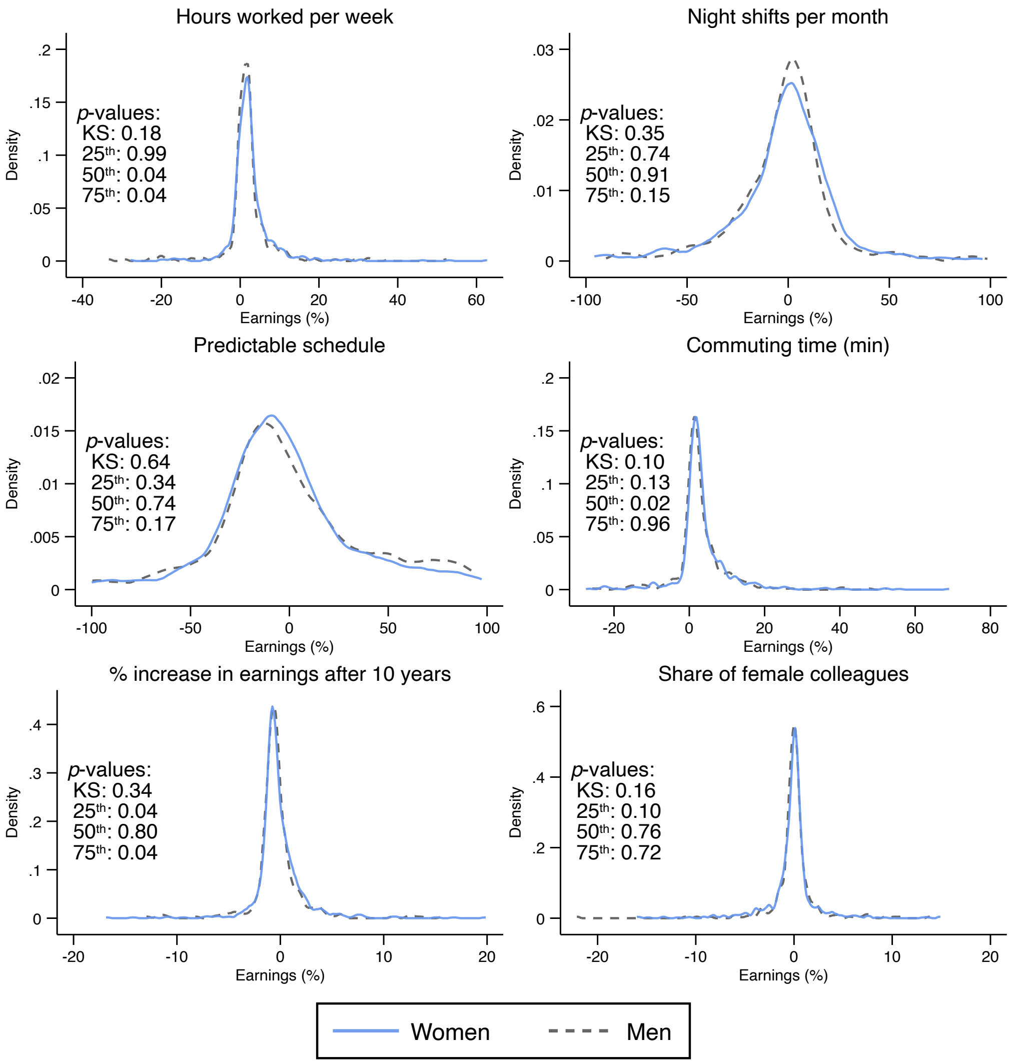
Although very informative, the hypothetical job choice exercise by design only allows us to measure taste for a limited set of characteristics. Could it be that this exercise failed to highlight important gender differences in taste for certain job characteristics? In this subsection, we look at the respondents' self-

[44]Note that given the specificity of the medical setting, we did expect a significant number of respondents to be indifferent to earnings.

[45]Note that we fix average earnings at the same level for men and women to ensure that gender differences in WTP reflect only differences in preferences, not earnings.

[46]In an effort to provide evidence on the quality of our preference estimates, Appendix E.3.4 shows the correlation that exists between the individual preference parameters estimated in this section and the characteristics of the medical specialties that are actually chosen by the survey respondents for their residency training.

Figure 4: Kernel Densities of WTP Estimates.



Notes: This figure displays the kernel densities of the percentage an individual needs to be compensated for a unit change in the job attribute or willingness to pay (WTP) computed using the relevant utility weight estimated from equation (4) by gender. p-values for Kolmogorov-Smirnov tests of the equality of the two distributions. We also conduct tests for gender differences in the 25^{th}, 50^{th} and 75^{th} percentiles using bootstrapped samples. WTP values above (below) 100 (-100) are not shown.

reported tastes for a wider set of job attributes. Specifically, we analyse a series of questions asking respondents to reveal how important a set of 13 characteristics are to them when choosing their specialty on a scale from 0 (not important at all) to 10 (extremely important).[47]

It is important to bear in mind that any potential discrepancies between the results of the discrete choice experiment and these self-reported measures could be due to the differences that exist between the two methods. The hypothetical job choice experiment is designed to estimate measures of individual preferences for job attributes that are free from potential confounding factors. In contrast, self-reported preferences for job characteristics are likely to be influenced by other correlating factors.

[47] Note that although French is a language with grammatical gender, all the questions and statements described below were written so as to apply to both genders, by using a gender-neutral phrasing.

For the sake of exposition, we reduce the dimensionality of this data by combining the characteristics as follows: *Willingness to help others* and *Contact with patients* are aggregated into *Patient care*; *Family-work balance*, *Workload* and *Schedule predictability* give rise to *Time flexibility*; and *Interest in the pathologies* and *Daily tasks* lead to *Job tasks*. We additionally decide to focus on *Income* and *Prestige* separately.

Figure 5 shows the mean of each of these indices separately by gender, while Appendix F Figure F1 replicates Figure 5 using the 13 original variables.[48] We also test the gender difference in means for each of these indices and report its significance level, and provide the percent difference between the female and male means for readability purposes. The figure complements the results of the hypothetical job choice experiment: women care more than men about having jobs that have a social component and involve more patient care (regular contact with patients, willingness to help others) and that have a flexible and predictable schedule (workload, family-work balance, schedule predictability). However, men put more importance than women on their job's income levels and prestige. Interestingly, the tasks required in a job (pathologies which are studied, presence or absence of surgery, of research, of emergencies) are similarly important for both genders.[49]

In addition, we asked respondents to select the most important driver of their specialty choice from the same list of job characteristics. We show their answers separately by gender in Appendix F Figure F2. For both men and women, the three most important factors behind their specialty choice are the work-life balance offered by the job, the interest they have in the specific pathologies, and having regular contact with patients. Women are significantly more likely than men to claim that their main consideration when choosing their specialty is the work-family balance, and to report the contact with patients as the main driver of their specialty choice.

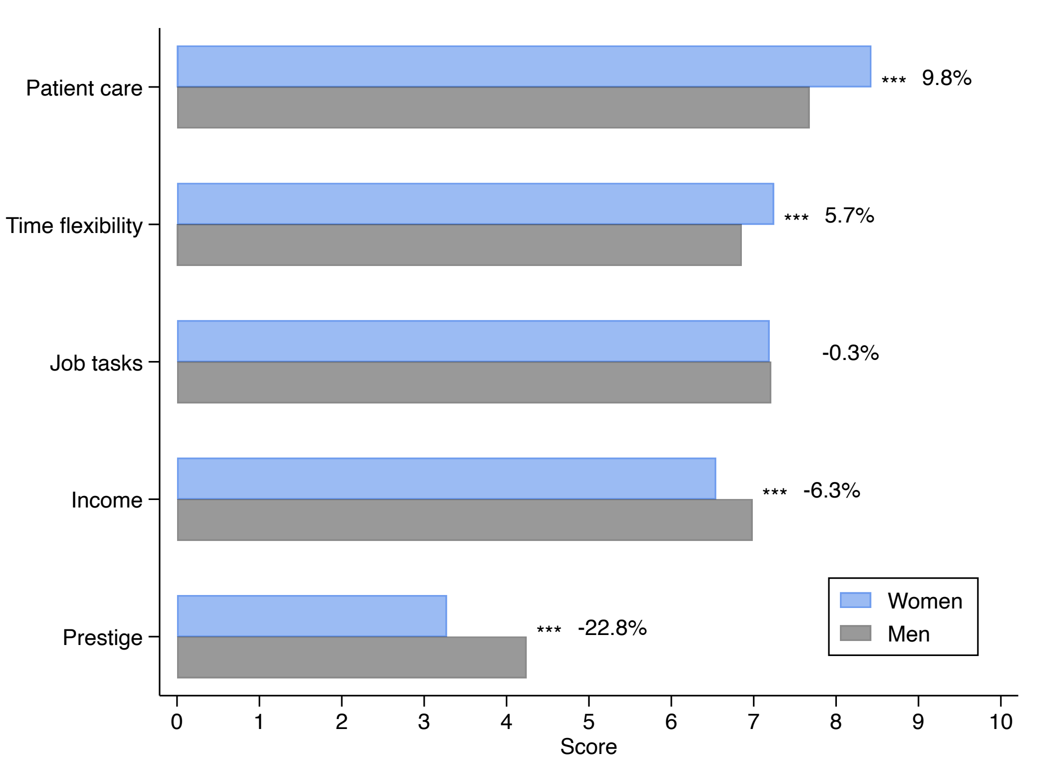
Finally, Appendix F.1 Figure F3 uses the respondents' answers to open-ended questions about the 'pros' and 'cons' of their preferred specialty to show that men and women, on average, differ in what they care about in a job. These open-ended answers also provide evidence of the similarity between the factors on which we surveyed respondents' self-reported tastes and the factors they mentioned when asked to openly state the 'pros' and 'cons' of their preferred specialty.

All in all, while our WTP estimates are remarkably similar across genders, our self-reported measures of preferences highlight some significant gender differences in tastes for job characteristics. While both men and women care about family-work balance and having control over their workload, these characteristics seem to be more important for women. Additionally, even though both men and women care about "softer" job characteristics such as care to patients, specific job tasks, and a friendly working environment, women care about these more than men. Finally, men care significantly more than women about income and

[48] Unlike in Section 5.2.3, we focus on gender differences in means for the sake of exposition, but looking at differences in distributions would imply a similar conclusion of important gender differences.

[49] It is also interesting to note that for both genders, the most important index in this list of job characteristics is patient care.

Figure 5: Self-Reported Valuation of Job Attributes by Gender.



Notes: This figure shows mean values for the five indices created based on the following survey question: *"At the beginning of this questionnaire, you said that you would like to specialise in specialty_preferred. On a scale from 0 to 10, how important was each of the following factors in your specialty choice?"*. Significance levels of the two-sample t-tests for the difference in means between men and women are reported using *** p<0.01, ** p<0.05, * p<0.1. A percent difference between the female and male means is also provided for each index.

prestige. As explained above, they way in which workplace preferences are elicited could explain the discrepancy between the conclusions of the discrete choice experiment and the self-reported measures. This could be the case if for instance, women do not dislike longer hours or night shifts *per se*, but dislike the types of tasks that are performed in the jobs in which they are required, more generally. An additional potential explanation is that the hypothetical job choice experiment requires more time and effort from respondents than the series of questions on self-reported preferences. In any case, the results from both methodologies suggest that a group of women have a stronger taste for time flexibility and patient care than men, while men put more importance on their job's prestige and income level than women. These gender differences are in line with those found in other studies, see Cortes and Pan (2018) for a summary of the findings in the literature. In the next section, we examine how these gender differences in workplace preferences impact medical specialty choice and the related gender earnings gap.

6 Preferences for Job Characteristics and Specialty Choice

In the previous section, we estimate individual-level valuations for certain workplace characteristics using a hypothetical job choice model and document how these compare to self-reported measures of taste for similar characteristics. In this section, we investigate how important these valuations are in explaining the observed gender-based occupational segregation and, more tentatively, the gender earnings gap. To do so, we group specialties according to whether they are predominantly preferred by women or by men over the entire period (2010-2022) based on Figure 2. Cardiology, anesthesiology, surgery and radiology are classified as *male-preferred*, while dermatology, pediatrics, general practice and gynecology are referred to as *female-preferred*. We then assess the effect of preferences for job characteristics on the gender gap in sorting into these female- and male-preferred specialties using two indicator variables that indicate whether an individual has chosen a specialty that belongs to one of these groups.

The role of workplace preferences in explaining specialty sorting To understand how much of the gender gap in the probability of specialising in male- and female-preferred specialties can be explained by differences in workplace preferences, we integrate our measures of preferences estimated in the previous section into a linear probability model of specialty choice. Table 4 shows the different specifications of that model. We first control for preferences for "hard" job characteristics only (income and time requirements) using in turn the β preference parameters estimated from equation (4) (column (2)) and the related self-reported measures (column (3)). Second, we additionally control for preferences for "softer" (self-reported) job characteristics, namely patient care, job tasks and prestige (columns (4) and (5)).

Finally, in order to control for one's knowledge and expectations of workplace characteristics, we use the respondents' answers to the survey questions asking them about their knowledge of the working conditions in one randomly-selected female-preferred specialty and one randomly-selected male-preferred specialty (column (6)). Specifically, we ask them to estimate the average earnings, the number of hours worked (including night shifts), the proportion of women, and proportion of self-employed physicians in these specialties. We adjust their answers for their self-reported confidence in the accuracy of each value provided. Overall, these specifications allow us to see how much of the gender gap in sorting remains after accounting for differences in preferences for job attributes and for potential differences in knowledge of and expectations on the specialties.

Table 4 displays the estimated gender gaps in the probability of self-selecting into a male-preferred specialty (Panel A) and into a female-preferred specialty (Panel B). Column (1) shows that after conditioning on rank deciles only, women are 11 percentage points less likely to self-select into male-preferred specialties and 22 percentage points more likely to self-select into female-preferred specialties. In column (2), we further control for the β parameters of workplace preferences estimated from equation (4). Controlling for these estimates only reduces the gender gaps in specialty choice by 0.2 percentage points in Panel A and 0.6 percentage points in Panel B, representing a 2 percent reduction in the residual gender gap in Panel A (-0.002/-0.11) and a 3 percent reduction in Panel B (0.006/0.216). This means that our "pure" measures of taste for hours worked, night shifts, schedule predictability, commuting time, earnings,

Table 4: Female-to-Male Gap in Specialty Choice.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Choosing a male-preferred specialty							
Female	-0.109***	-0.107***	-0.082***	-0.073***	-0.062***	-0.055**	-0.055**	-0.047*
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.026)
Wants kids							0.000	0.026
							(0.020)	(0.047)
Female x Wants kids								-0.035
								(0.051)
Mean dep. variable	0.156	0.156	0.156	0.156	0.156	0.156	0.156	0.156
Adj. R-squared	0.163	0.166	0.206	0.217	0.234	0.241	0.239	0.239
Observations	1,134	1,134	1,134	1,134	1,134	1,134	1,134	1,134
Exam rank deciles	✓	✓	✓	✓	✓	✓	✓	✓
Workplace preference controls:								
Hard job charac. (betas)	-	✓	-	✓	-	✓	✓	✓
Hard job charac. (self-reported)	-	-	✓	-	✓	-	-	-
Soft job charac. (self-reported)	-	-	-	✓	✓	✓	✓	✓
Workplace knowledge	-	-	-	-	-	✓	✓	✓
	Panel B: Choosing a female-preferred specialty							
Female	0.216***	0.210***	0.186***	0.164***	0.152***	0.144***	0.139***	0.132***
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.028)	(0.029)	(0.032)
Wants kids							0.071**	0.047
							(0.029)	(0.055)
Female x Wants kids								0.034
								(0.064)
Mean dep. variable	0.522	0.522	0.522	0.522	0.522	0.522	0.522	0.522
Adj. R-squared	0.114	0.124	0.144	0.200	0.207	0.266	0.269	0.269
Observations	1,134	1,134	1,134	1,134	1,134	1,134	1,134	1,134
Exam rank deciles	✓	✓	✓	✓	✓	✓	✓	✓
Workplace preference controls:								
Hard job charac. (betas)	-	✓	-	✓	-	✓	✓	✓
Hard job charac. (self-reported)	-	-	✓	-	✓	-	-	-
Soft job charac. (self-reported)	-	-	-	✓	✓	✓	✓	✓
Workplace knowledge	-	-	-	-	-	✓	✓	✓

Notes: OLS estimates of the gender difference in the probability of selecting Cardiology, Anesthesiology, Surgery or Radiology (Panel A) and Dermatology, General Practice, Gynecology or Pediatrics (Panel B). Refer to Section 5.1 for details on the data. All specifications control for exam rank deciles. Workplace preferences controls are defined as follows: *Hard job charac. (betas)* includes the estimated preferences parameters (β's) from the hypothetical job choice model; *Hard job charac. (self-reported)* includes the self-reported preferences measures for income and time flexibility; *Soft job charac. (self-reported)* includes the indices for patients care, job tasks, and prestige. *Workplace knowledge* includes the respondents' expectations of earnings, hours worked, share of women and share of self-employed physicians in the specialty adjusted for confidence. Wants to have kids takes value 1 for those who want to have kids within the next 5 years and 0 for those who answered no. Respondents who already have kids are not included. Block bootstrap standard errors in parentheses. ***, **, * denote significance at 1%, 5%, and 10% levels, respectively.

earnings growth and share of female colleagues contribute only marginally to the gender-based occupational segregation that we observe in the data. This result is not surprising, as we have shown in Section 5.2.3 that these preferences estimates are remarkably similar across genders. In column (3), instead of the β preference parameters, we use the respondents' self-reported preferences for earnings, workload, schedule predictability, and family-work balance to control for taste for time flexibility and earnings. The resulting reductions in gender gaps (and increases in adjusted R-squared) are larger than those observed in column (2), but they are still rather limited: the residual gender gap is reduced by 2.5 percentage points in Panel A and 2.4 percentage points in Panel B, representing a 25 percent lower probability of self-selecting into a male-preferred specialty (Panel A) and a 14 percent lower probability of self-selecting into a female-preferred specialty (Panel B). These differences in effect sizes are likely due to the different methodologies used to obtain these individual measures of taste, as highlighted in Subsection 5.3.

Column (4) builds on column (2) by adding the self-reported preference parameters that are not related to time flexibility and earnings, namely interest in the pathologies, daily tasks, contact with patients, prestige, and willingness to help others. Interestingly, taste for these "softer" job characteristics seem to contribute much more to reducing the gender gap in sorting than taste for the "harder" ones that are earnings and time flexibility. Adding them to the specification significantly reduces both the gender gap in the probability of choosing a male-preferred specialty (by an additional 33 percent) and a female-preferred specialty (by an additional 24 percent), and substantially increases the adjusted R-squared in both panels. Column (5), in which we control for self-reported preferences for both the hard and soft job characteristics leads to a similar conclusions: the latter explain more of the gender gap in sorting than the former. Our results are consistent with those of studies analysing gender-based occupational segregation finding that women tend to have stronger preferences than men for jobs with a higher people component (Lordan and Pischke, 2022; Cortes and Pan, 2018). Similarly, studies on the determinants of the gender wage gap have found that preferences for people/family versus money/work can account for at least as much or even more than other traditionally relevant human capital factors in explaining the residual gender wage gap (Fortin, 2008; Grove, Hussey and Jetter, 2011; Cortes and Pan, 2018). Column (6) additionally controls for individual knowledge of mean earnings, number of hours worked (including night shifts), share of women, and share of self-employed physicians in the randomly-drawn specialties. It shows that individual-level expectations about job characteristics helps explain the gender gap in specialty sorting, consistent with previous literature on the role of information on occupational choices (Wiswall and Zafar, 2018) and on college major (Zafar, 2013; Wiswall and Zafar, 2015).

The role of preference for fertility in explaining specialty sorting Finally, we examine how much of the difference in sorting can be explained by differences in preferences for future fertility by adding an indicator variable taking the value 1 if the respondent reports wanting to have children within the next five years and 0 otherwise. Column (7) shows that, on average, controlling for the willingness to have children in the near future does not further reduce the sorting gap into male-preferred specialties, but does so by 3.5 percent when it comes to sorting into female-preferred specialties. In other words, wanting to have children in the near future is associated with a higher probability of choosing a female-preferred specialty. Column (8) further allows us to investigate whether preferences for future fertility play a different role for men and women. We find that the gender difference in sorting is smaller between women and men who do not want to have children in the next five years (4.7 percent for male-preferred specialty and 13.2 percent for female-preferred specialty) than between men and women who do want to have children in the next five years (5.6 for male-preferred specialty and 21.3 for female-preferred specialty). In addition, although the estimates on the interaction terms are not precisely estimated, they suggest that while wanting children in the near future is associated with a higher probability of self-selecting into a female-preferred specialty for both men and women, it is associated with a lower probability of self-selecting into male-preferred specialty for women but not for men. This finding point towards an anticipation effect of fertility on the career choices of young women, consistent with the structural estimates obtained in Adda, Dustmann and Stevens (2017). It highlights the costs of fertility incurred by women well before the birth of their first child, which is in line with the literature on the relationship between women's fertility choices and gender gaps in labour market outcomes (Sasser, 2005; Wang and Sweetman, 2013; Angelov, Johansson and Lindahl, 2016; Kleven, Landais and Søgaard, 2019; Kleven et al., 2019; Wasserman, 2022, 2023). Overall, we find that preferences for and knowledge of workplace attributes, together with preferences for future fertility, explain 50 and 64 percent of the residual gender gap in sorting into male- and female-preferred specialties, respectively.

The role of gender differences in workplace preferences in explaining the gender earnings gap In the last part of this section, we examine the extent to which gender differences in job preferences explain the gender earnings gap. We have shown that the lion's share of the gender pay gap among physicians can be explained by speciality choice (Section 3), and that workplace preferences affect specialty choices. To quantify how workplace preferences affect the gender earnings gap through the specialty choice channel, we follow Wiswall and Zafar (2018) and implement an exercise that measures the gender earnings gap in a hypothetical scenario in which the distribution of women's workplace preferences is shifted by the median gender difference to have the male median.[50]

We do so by estimating a non-linear specialty choice model with the same control variables as in column (4) of Table 4, and predicting the probability that women choose male-preferred specialties (as opposed to female-preferred specialties) after shifting their workplace preferences by the median gender difference. This exercise relies on the survey questions eliciting individual estimates of expected earnings[51] in a female-preferred and a male-preferred randomly-selected specialty.[52] We then use each female respondent's predicted probability of choosing a male-preferred specialty to weight her self-reported expected earnings in each of the two specialty groups (female- and male-preferred). Note that we keep women's relative performance in the NRE fixed, even though it could be affected by a shift in workplace preferences.

We find that shifting women's workplace preferences to have the male median increases women's probability of choosing a male-preferred specialty by 15 percent, thereby reducing the expected gender

[50]Shifting women's workplace preferences by the median male-female difference allows us to preserve the heterogeneity in women's preferences.

[51]Note that unlike the analysis carried out in Section 3, the current analysis relies on *earnings* data. We made this decision for simplicity's sake, on the basis that it is difficult for medical students to give an estimate of how much they expect to earn on average, let alone how much they expect to earn per hour or per procedure.

[52]As a result, unlike the specialty choice model specified above, this exercise focuses on the decision to choose a male-preferred specialty as opposed to a female-preferred specialty, as we did not elicit knowledge of the working conditions in the remaining "neutral" specialties.

earnings gap between the male- and female-preferred specialties by 1.6 percent. In particular, the gender gap is reduced by 0.5 percentage points from a baseline predicted gap of 31.2 percent. Our results support the theory of compensating differentials, whereby the observed gender earnings gap could be partly due to women sorting into medical specialties with more favourable job attributes that pay less in return. Given our limited and rather crude measures of expected earnings in female- and male-preferred specialties, we think that our estimate of the reduction is likely to be a lower bound on the importance of workplace preferences in determining the gender earnings gap in the medical profession.

7 Conclusion

In this paper, we analyse gender differences in early career choices in a context in which the traditional explanations for gender-based occupational segregation, namely human capital investment and discrimination, are by construction not present: the French centralised medical specialty selection procedure. This allocation mechanism of medical students to residency positions allows us to focus on the role of supply-side factors in explaining gender-based occupational segregation and, to a certain extent, its implications for the gender pay gap in the profession.

To motivate our analysis of gender-based segregation in medical specialties, we start by showing that specialty of practice is the most important observable determinant of the gender earnings gap in the medical profession. We then use unique administrative data on the French centralised medical residency selection mechanism to show that, conditional on facing virtually equivalent medical residency choice sets, men and women make drastically different specialty choices. Moreover, we show that this result holds at the top of the exam performance distribution, where candidates face no external constraints on their decisions. The specialty choices stemming from this true preference revealing mechanism result in women being more likely than men to self-select into occupations with lower expected earnings, allow for more time flexibility, and are less competitive than those chosen by men. We argue that such a finding in this particular setting strongly suggests that workplace preferences play an important role in explaining the observed gender differences in early career choices.

To better understand the workplace preferences that lead men and women to sort into occupations with such different job characteristics, we turn to the field and survey a sample of French sixth year medical student in 2022, before they take the National Ranking Examinations, about their occupational aspirations, their expected exam performance and the drivers of their choices. Using a hypothetical job choice framework, we robustly estimate individual preferences for certain job attributes. We obtain willingness-to-pay estimates for these attributes that are remarkably similar across genders. Although preferences for the number of hours worked show some gender differences at the tail, suggesting the existence of a group of women with particularly strong tastes for time flexibility who are willing to give up large shares of monthly earnings to avoid working longer hours, we find no evidence of strong gender differences that could fully explain the gender gap in sorting documented above.

We complement these results with collected self-reported measures of preferences for job characteristics, which allows us to identify some further gender differences in preferences for job attributes that are more in line with the existing literature (Cortes and Pan, 2018). While both men and women care about family-work balance and having control over their workload, these characteristics seem to be more important for women. Additionally, although both men and women care about "soft" job characteristics such as care to patients, specific job tasks, and a friendly working environment, women do care more than men about patient care. Finally, men care significantly more than women about income and prestige.

We then combine estimated preferences for workplace attributes, knowledge of workplace attributes and preferences for future fertility to explore how much of the gender gap in specialty sorting they can explain. We show that while "hard" job characteristics such as earnings and time constraints matter for the gender gap in sorting, "soft" characteristics such as daily tasks, contact with patients, and willingness to help others play a substantially larger role in reducing the gap. Knowledge of workplace attributes is also found to be a relevant factor in explaining the differential sorting across specialties. Finally, we find suggestive evidence of an anticipation effect of fertility on the career choices of young women but not of men.

A potential area for future research is to investigate the role of social and environmental factors, such as expectations of future discrimination and social norms, in explaining gender differences in early career choices. Future research could also be devoted to the further understanding of the drivers behind the gender differences in specialisation in sub-fields of certain occupations (Goldin and Katz, 2016; Azmat and Ferrer, 2017).

References

Adda, Jérôme, Christian Dustmann, and Katrien Stevens. 2017. "The Career Costs of Children." *Journal of Political Economy*, 125(2): 293–337.

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity*." *The Quarterly Journal of Economics*, 115(3): 715–753.

Altonji, Joseph, and Rebecca Blank. 1999. "Race and gender in the labor market." In *Handbook of Labor Economics*. Vol. 3, Part C. Ashenfelter, O. and Card, D. ed., Chapter 48, 3143–3259. Elsevier.

Angelov, Nikolay, Per Johansson, and Erica Lindahl. 2016. "Parenthood and the Gender Gap in Pay." *Journal of Labor Economics*, 34(3): 545–579.

Azmat, Ghazala, and Barbara Petrongolo. 2014. "Gender and the labor market: What have we learned from field and lab experiments?" *Labour Economics*, 30(C): 32–40.

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

Bertrand, Marianne. 2011. "New Perspectives on Gender." In *Handbook of Labor Economics*. Vol. 4, 1543–1590. Elsevier.

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz. 2010. "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors." *American Economic Journal: Applied Economics*, 2(3): 228–55.

Biasi, Barbara, and Heather Sarsons. 2020. "Flexible Wages, Bargaining, and the Gender Gap." National Bureau of Economic Research Working Paper 27894.

Blass, Asher A., Saul Lach, and Charles F. Manski. 2010. "Using Elicited Choice Probabilities To Estimate Random Utility Models: Preferences For Electricity Reliability." *International Economic Review*, 51(2): 421–440.

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

Bolotnyy, Valentin, and Natalia Emanuel. 2022. "Why Do Women Earn Less Than Men? Evidence from Bus and Train Operators." *Journal of Labor Economics*, 40(2): 283–323.

Buser, Thomas, and Huaiping Yuan. 2019. "Do Women Give Up Competing More Easily? Evidence from the Lab and the Dutch Math Olympiad." *American Economic Journal: Applied Economics*, 11(3): 225–52.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. 2014. "Gender, Competitiveness, and Career Choices." *The Quarterly Journal of Economics*, 129(3): 1409–1447.

Charles, Kerwin Kofi, Jonathan Guryan, and Jessica Pan. 2018. "The Effects of Sexism on American Women: The Role of Norms vs. Discrimination." National Bureau of Economic Research, Inc NBER Working Papers 24904.

Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. "oTree—An open-source platform for laboratory, online, and field experiments." *Journal of Behavioral and Experimental Finance*, 9: 88–97.

Conseil national de l'ordre des médecins. 2016. "Atlas de la démographie médicale en France."

Cook, Cody, Rebecca Diamond, Jonathan V Hall, John A List, and Paul Oyer. 2021. "The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers." *The Review of Economic Studies*, 88(5): 2210–2238.

Cortes, Patricia, and Jessica Pan. 2018. "Occupation and Gender." *The Oxford Handbook of Women and the Economy*, , ed. Susan L. Averett, Laura M. Argys and Saul D. Hoffman . online ed., Chapter 18. Oxford Handbooks.

Coudin, Elise, Anne Pla, and Anne-Laure Samson. 2015. "GP Responses to Price Regulation: Evidence from a French Nationwide Reform." *Health Economics*, 24(9): 1118–1130.

DeLeire, Thomas, and Helen Levy. 2004. "Worker Sorting and the Risk of Death on the Job." *Journal of Labor Economics*, 22(4): 925–954.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde; Jürgen Schupp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Consequences." *Journal of the European Economic Association*, 9.

Dorsey, E Ray, David Jarjoura, and Gregory W Rutecki. 2003. "Influence of controllable lifestyle on recent trends in specialty choice by US medical students." *Jama*, 290(9): 1173–1178.

Díez-Rituerto, Marina, Javier Gardeazabal, Nagore Iriberri, and Pedro Rey-Biel. 2022. "Gender Gaps in Access to Medical Intern Positions: The Role of Competition." *CEPR Press Discussion Paper No. 17702.*

Eriksson, Tor, and Nicolai Kristensen. 2014. "Wages or Fringes? Some Evidence on Trade-Offs and Sorting." *Journal of Labor Economics*, 32(4): 899–928.

Fadlon, Itzik, Frederik Plesner Lyngse, and Torben Heien Nielsen. 2020. "Early Career, Life-Cycle Choices, and Gender." National Bureau of Economic Research Working Paper 28245.

Flory, Jeffrey A., Andreas Leibbrandt, and John A. List. 2014. "Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions." *The Review of Economic Studies*, 82(1): 122–155.

Fluchtmann, Jonas, Anita M Glenny, Nikolaj Harmon, and Jonas Maibom. 2020. "Do men and women apply for the same jobs?" 94.

Fortin, Nicole. 2008. "The Gender Wage Gap Among Young Adults in the United States: The Importance of Money Versus People." *Journal of Human Resources*, 43: 884–918.

Gagné, Robert, and Pierre Thomas Léger. 2005. "Determinants of physicians' decisions to specialize." *Health Economics*, 14(7): 721–735.

Gale, David, and Lloyd S. Shapley. 1962. "College Admissions and the Stability of Marriage." *The American Mathematical Monthly*, 69: 9–15.

Goldin, Claudia. 2002. "A Pollution Theory of Discrimination: Male and Female Occupations and Earnings." NBER Working Paper no. 8985.

Goldin, Claudia. 2004. "The Long Road to the Fast Track: Career and Family." *The Annals of the American Academy of Political and Social Science*, 596(1): 20–35.

Goldin, Claudia. 2006. "The Quiet Revolution That Transformed Women's Employment, Education, and Family." *American Economic Review*, 96(2): 1–21.

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

Goldin, Claudia. 2015. ""How to Achieve Gender Equality"." *Milken Institute Review*, July(Q3): 24–33.

Goldin, Claudia, and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians." *American Economic Review*, 90(4): 62.

Goldin, Claudia, and Lawrence F. Katz. 2016. "A Most Egalitarian Profession: Pharmacy and the Evolution of a Family Friendly Occupation." *Journal of Labor Economics*, 34(3): 705–745.

Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the Gender Gap in College." *Journal of Economic Perspectives*, 20: 133–156.

Golfouse, Anny, and Bunna Pheng. 2015. “Les épreuves classantes nationales (ECN) donnant accès au 3ème cycle des études médicales.” Observatoire National de la Démographie des Professions de Santé.

Grove, Wayne A., Andrew Hussey, and Michael Jetter. 2011. “The Gender Pay Gap Beyond Human Capital: Heterogeneity in Noncognitive Skills and in Labor Market Tastes.” *Journal of Human Resources*, 46(4): 827–874.

Harris, Jeffrey E, Beatriz G Lopez-Valcarcel, Patricia Barber, and Vicente Ortún. 2017. “Allocation of residency training positions in Spain: contextual effects on specialty preferences.” *Health Economics*, 26(3): 371–386.

Ketel, Nadine, Edwin Leuven, Hessel Oosterbeek, and Bas van der Klaauw. 2016. “The Returns to Medical School: Evidence from Admission Lotteries.” *American Economic Journal: Applied Economics*, 8(2): 225–54.

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

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and Josef Zweimuller. 2019. “Child Penalties across Countries: Evidence and Explanations.” *AEA Papers and Proceedings*, 109: 122–26.

Ku, Manwai C. 2011. “When Does Gender Matter?: Gender Differences in Specialty Choice Among Physicians.” *Work and Occupations*, 38(2): 221–262.

Le Barbanchon, Thomas, Roland Rathelot, and Alexandra Roulet. 2020. “Gender Differences in Job Search: Trading off Commute against Wage*.” *The Quarterly Journal of Economics*, 136(1): 381–426.

Lordan, Grace, and Jörn-Steffen Pischke. 2022. “Does Rosie Like Riveting? Male and Female Occupational Choices.” *Economica*, 89(353): 110–130.

Maestas, Nicole, Kathleen J. Mullen, David Powell, Till von Wachter, and Jeffrey B. Wenger. 2018. “The Value of Working Conditions in the United States and Implications for the Structure of Wages.” Publisher: RAND Corporation.

Mas, Alexandre, and Amanda Pallais. 2017. “Valuing Alternative Work Arrangements.” *American Economic Review*, 107(12): 3722–3759.

Mcfadden, Daniel, and Kenneth Train. 2000. “Mixed MNL models for discrete response.” *J. Appl. Econ.*, 24.

Neumark, David. 2018. "Experimental Research on Labor Market Discrimination." *Journal of Economic Literature*, 56(3): 799–866.

Neyman, J., and Elizabeth L. Scott. 1948. "Consistent Estimates Based on Partially Consistent Observations." *Econometrica*, 16(1): 1–32.

Nicholson, Sean. 2002. "Physician specialty choice under uncertainty." *Journal of Labor Economics*, 20(4): 816–847.

Porter, Catherine, and Danila Serra. 2020. "Gender Differences in the Choice of Major: The Importance of Female Role Models." *American Economic Journal: Applied Economics*, 12(3): 226–54.

Reuben, Ernesto, Matthew Wiswall, and Basit Zafar. 2017. "Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender." *Economic Journal*, 127(604): 2153–2186.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2015. "Taste for Competition and the Gender Gap Among Young Business Professionals." National Bureau of Economic Research Working Paper 21695.

Riach, P. A., and J. Rich. 2002. "Field Experiments of Discrimination in the Market Place*." *The Economic Journal*, 112(483): F480–F518.

Rich, Judith. 2014. "What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted Since 2000." *SSRN Electronic Journal*.

Sasser, Alicia C. 2005. "Gender Differences in Physician Pay: Tradeoffs Between Career and Family." *Journal of Human Resources*, XL(2): 477–504.

Sivey, Peter, Anthony Scott, Julia Witt, Catherine Joyce, and John Humphreys. 2012. "Junior doctors' preferences for specialty choice." *Journal of Health Economics*, 31(6): 813–823.

Sockin, Jason. 2021. "Show Me the Amenity: Are Higher-Paying Firms Better All Around?" *SSRN Electronic Journal*.

Thornton, James, and Fred Esposto. 2003. "How important are economic factors in choice of medical specialty?" *Health Economics*, 12(1): 67–73.

Wang, Chao, and Arthur Sweetman. 2013. "Gender, family status and physician labour supply." *Social Science & Medicine*, 94: 17–25.

Wasserman, Melanie. 2022. "Hours Constraints, Occupational Choice, and Gender: Evidence from Medical Residents." *The Review of Economic Studies*.

Wasserman, Melanie. 2023. "Do Long Work Hours Constrain Women's Fertility Choice?"

Wiswall, Matthew, and Basit Zafar. 2015. "Determinants of College Major Choice: Identification using an Information Experiment." *The Review of Economic Studies*, 82(2 (291)): 791–824. Publisher: [Oxford University Press, The Review of Economic Studies, Ltd.].

Wiswall, Matthew, and Basit Zafar. 2018. "Preference for the Workplace, Investment in Human Capital, and Gender." *The Quarterly Journal of Economics*, 133(1): 457–507.

Zafar, Basit. 2013. "College Major Choice and the Gender Gap." *Journal of Human Resources*, 48(3): 545–595.

Appendix

Appendix A Supplementary Material on the Data

A.1 The National Ranking Examinations

Figure A1: Subdivisions

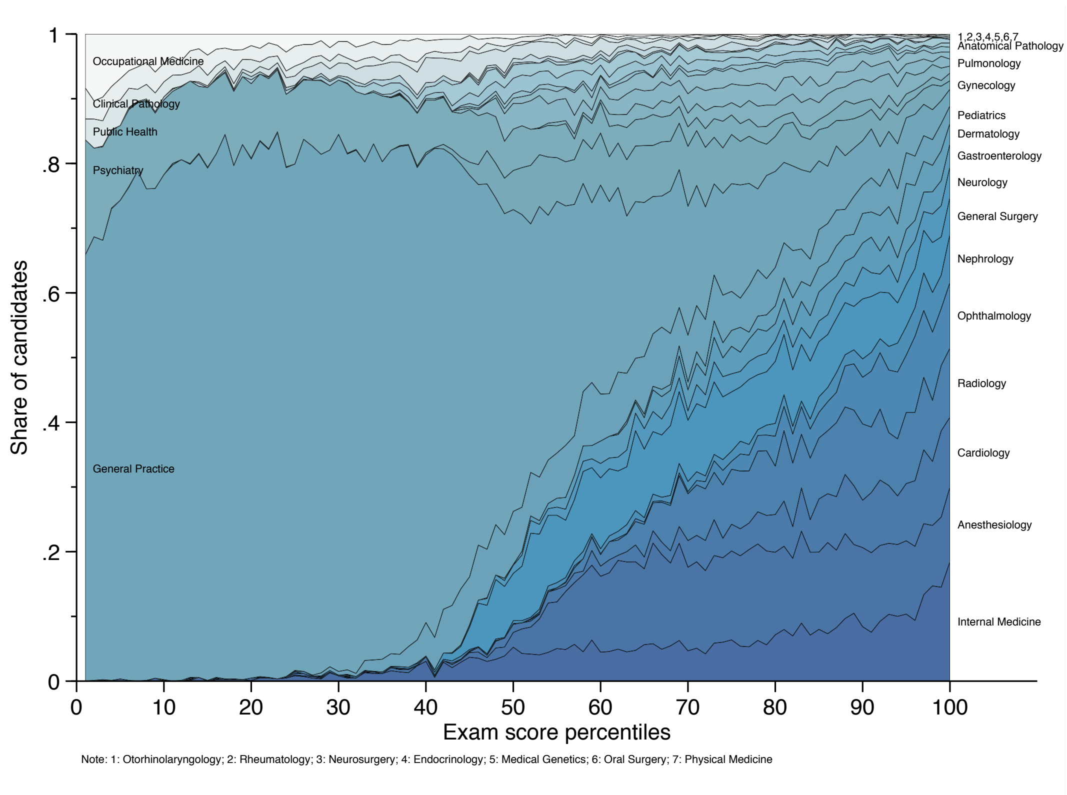
Notes: This figure shows the 26 subdivisions of mainland France. Subdivisions are geographical areas comprising one or several teaching hospitals. They are the geographical unit of reference in the context of medical studies in France. There are two additional subdivisions in overseas France: Antilles-Guyane and Océan Indien.

Table A1: Descriptive Statistics on the National Ranking Examinations (2010-2022), Before and After Sample Selection.

	All	Women	Men	Diff.	Obs.
Nb. vacancies	8,188.5				
	(623.6)				
Nb. takers	8,456.0	5,000.1	3,455.9	1,377.6	
	(722.8)	(507.1)	(310.8)		
			Before sample selection		
% Women	0.590				108,657
	(0.492)				
Age	25.2	25.0	25.4	-0.4***	72,772
	(2.0)	(1.8)	(2.2)		
Rank	4,202.9	4,230.8	4,162.9	67.9***	108,657
	(2,451.6)	(2,379.8)	(2,550.8)		
% Available positions	0.453	0.446	0.464	-0.018***	108,657
	(0.310)	(0.302)	(0.321)		
			After sample selection		
% Women	0.595				99,076
	(0.491)				
Age	25.2	25.0	25.4	-0.4***	66,146
	(2.0)	(1.8)	(2.2)		
Rank	4,074.9	4,113.7	4,018.0	95.7***	99,076
	(2,449.9)	(2,370.0)	(2,561.8)		
% Available positions	0.470	0.461	0.485	-0.024***	99,076
	(0.312)	(0.304)	(0.323)		

Notes: This table reports descriptive statistics on the National Ranking Examinations between 2010 and 2022. *Nb. vacancies* and *Nb. takers* respectively report the number of vacancies that are offered on a given year on average, and the number of candidates taking the exam on a given year on average. The table reports the average value for the whole sample (1), the sample of women (2) and the sample of men (3), as well as differences in means between women and men (4) and the number of observations (5). Additionally, standard deviations for the means are reported in parenthesis, and significance levels of the two-sample t-tests for the difference in means between women and men are reported using *** p<0.01, ** p<0.05, * p<0.1.

Figure A2: Share of Candidates Selecting each Specialty.



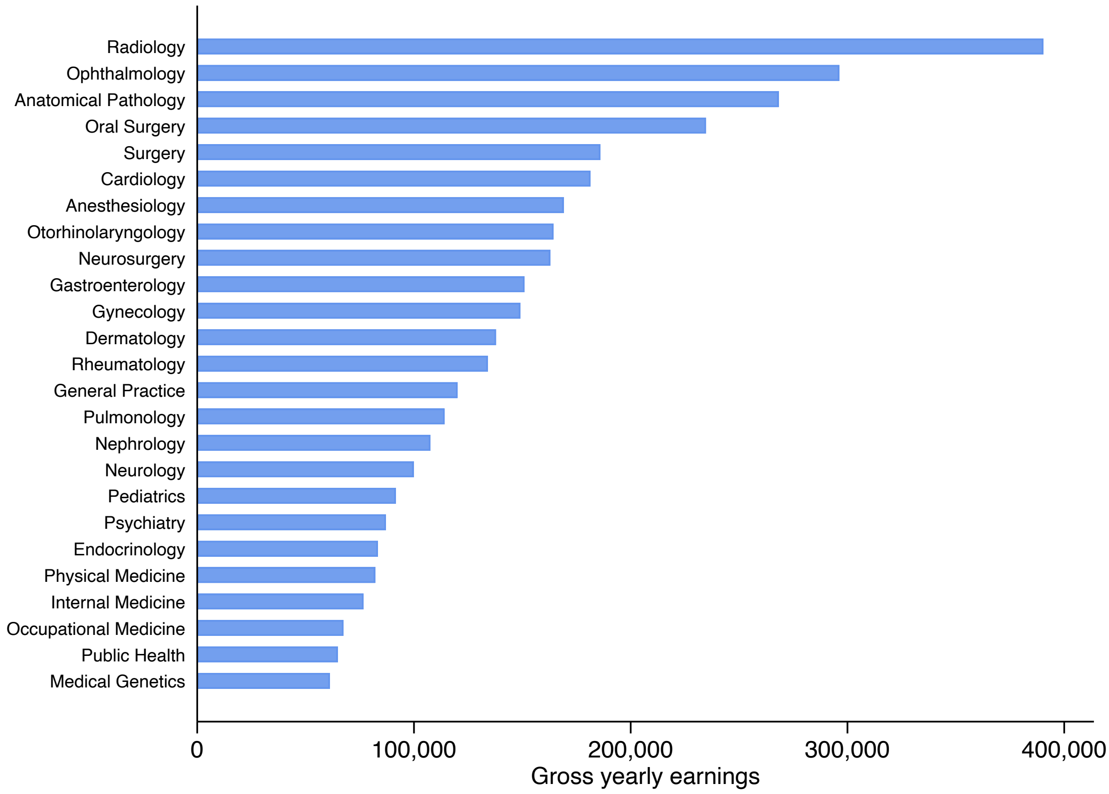
Notes: This figure plots the share of candidates selecting each specialty in each percentile of the exam score distribution.

A.2 Specialty Characteristics

In order to characterise the medical specialties that men and women choose from, we gather data on measurable job attributes that matter in the occupational decision. In this section, we describe the different data sources from which we obtain these pecuniary and non-pecuniary attributes, and provide some descriptive statistics for each of them.

Expected Earnings Table A2 provides descriptive statistics on the self-employed physicians whose data is described in Section 2.2. Figure A3 plots our proxy of gross yearly expected earnings by specialty described in Section 2.2. It shows that there is significant variation in expected earnings across specialties, our proxy ranging from only slightly more than $60,000$ euros in public health and medical genetics to close to $390,000$ euros in radiology.

Figure A3: Expected Yearly Earnings by Specialty.



Notes: This figure shows the yearly gross earnings that physicians can expect to earn in each specialty on average. This measure combines aggregated data on the earnings of private physicians provided by the CNAMTS and hospital salary grids collected from the Official Journal of the French Republic, which are averaged and weighted by population using data on demographics provided by the DREES.

Time Requirements Turning to non-pecuniary characteristics, one of the most prominent job attributes for which men and women are likely to differ in their preferences is their work's time requirements. In the medical context, two measures that are particularly relevant are the duration of the work week and the number of night shifts. Like salaries, hours worked and night shifts are subject to regulations: medical

Table A2: Descriptive Statistics Self-employed Phyisicians.

	Sector 1			Sector 2		
	Women	Men	Diff. (p-value)	Women	Men	Diff. (p-value)
Yearly gross earnings	130,292	184,919	-54,628 (0.00)	158,312	241,468	-83,157 (0.00)
Earnings per procedure	49.21	44.66	4.55 (0.14)	59.18	85.32	-26.14 (0.00)
per consultation	17.28	20.14	-2.86 (0.01)	22.37	24.39	-2.02 (0.00)
per intervention	158.65	191.63	-32.98 (0.01)	102.75	210.05	-107.30 (0.00)
Share earnings from self-employment	0.91	0.89	0.02 (0.00)	0.86	0.84	0.02 (0.00)
Performed medical procedures	4166.73	5556.66	-1389.93 (0.00)	3079.34	3764.30	-684.96 (0.00)
share consultations	0.81	0.79	0.03 (0.00)	0.76	0.68	0.08 (0.00)
share interventions	0.19	0.21	-0.03 (0.00)	0.24	0.32	-0.08 (0.00)
Number of different patients	1,849	2,091	-242 (0.00)	1,726	1,856	-131 (0.00)
Years since specialisation	19.57	23.75	-4.18 (0.00)	23.09	24.14	-1.06 (0.00)
Years in self-employment	15.93	20.78	-4.85 (0.00)	18.74	20.03	-1.28 (0.00)
Age						
Below 45	0.30	0.15	0.15 (0.00)	0.21	0.17	0.04 (0.00)
40-49	0.19	0.15	0.04 (0.00)	0.13	0.12	0.01 (0.00)
50-54	0.21	0.22	-0.00 (0.02)	0.20	0.18	0.02 (0.00)
55-60	0.18	0.25	-0.07 (0.00)	0.24	0.24	0.00 (0.40)
Above 60	0.12	0.24	-0.12 (0.00)	0.21	0.29	-0.07 (0.00)
Civil status						
Single	0.21	0.07	0.14 (0.00)	0.17	0.06	0.11 (0.00)
Divorced	0.15	0.10	0.05 (0.00)	0.18	0.12	0.06 (0.00)
Married/civil partnership	0.62	0.82	-0.19 (0.00)	0.63	0.81	-0.18 (0.00)
Widow	0.02	0.01	0.01 (0.00)	0.02	0.01	0.01 (0.00)
Number of children	1.29	1.25	0.04 (0.00)	1.12	1.27	-0.15 (0.00)
Child age < 5	0.11	0.08	0.03 (0.00)	0.09	0.11	-0.02 (0.00)
Share by medical specialty						
Anesthesiology	0.02	0.03	-0.01 (0.00)	0.03	0.04	-0.02 (0.00)
Cardiology	0.02	0.05	-0.03 (0.00)	0.02	0.03	-0.02 (0.00)
Dermatology	0.05	0.01	0.04 (0.00)	0.10	0.03	0.07 (0.00)
Gastroenterology	0.01	0.02	-0.01 (0.00)	0.02	0.03	-0.01 (0.00)
General Practice	0.60	0.65	-0.05 (0.00)	0.22	0.24	-0.02 (0.00)
Gynecology	0.06	0.02	0.04 (0.00)	0.17	0.08	0.10 (0.00)
Ophthalmology	0.04	0.02	0.02 (0.00)	0.11	0.08	0.03 (0.00)
Oral Surgery	0.00	0.01	-0.00 (0.00)	0.00	0.02	-0.01 (0.00)
Others	0.04	0.03	0.01 (0.00)	0.07	0.04	0.03 (0.00)
Otorhinolaryngology	0.01	0.01	-0.01 (0.00)	0.02	0.06	-0.04 (0.00)
Pediatrics	0.04	0.01	0.03 (0.00)	0.05	0.02	0.03 (0.00)
Psy And Neuropsychiatry	0.07	0.05	0.02 (0.00)	0.08	0.06	0.03 (0.00)
Pulmonology	0.01	0.01	-0.00 (0.00)	0.01	0.01	-0.00 (0.70)
Radiology	0.04	0.06	-0.02 (0.00)	0.03	0.02	0.00 (0.72)
Rheumatology	0.01	0.01	0.00 (0.09)	0.03	0.03	0.01 (0.00)
Surgery	0.00	0.02	-0.02 (0.00)	0.03	0.21	-0.18 (0.00)
Number of physicians	93,984	218,169	312,153	28,400	72,847	101,247

Notes: Mean values of selected variables from administrative data on all physicians in France with positive earnings from self-employment in the years 2005, 2008, 2011 and 2014.

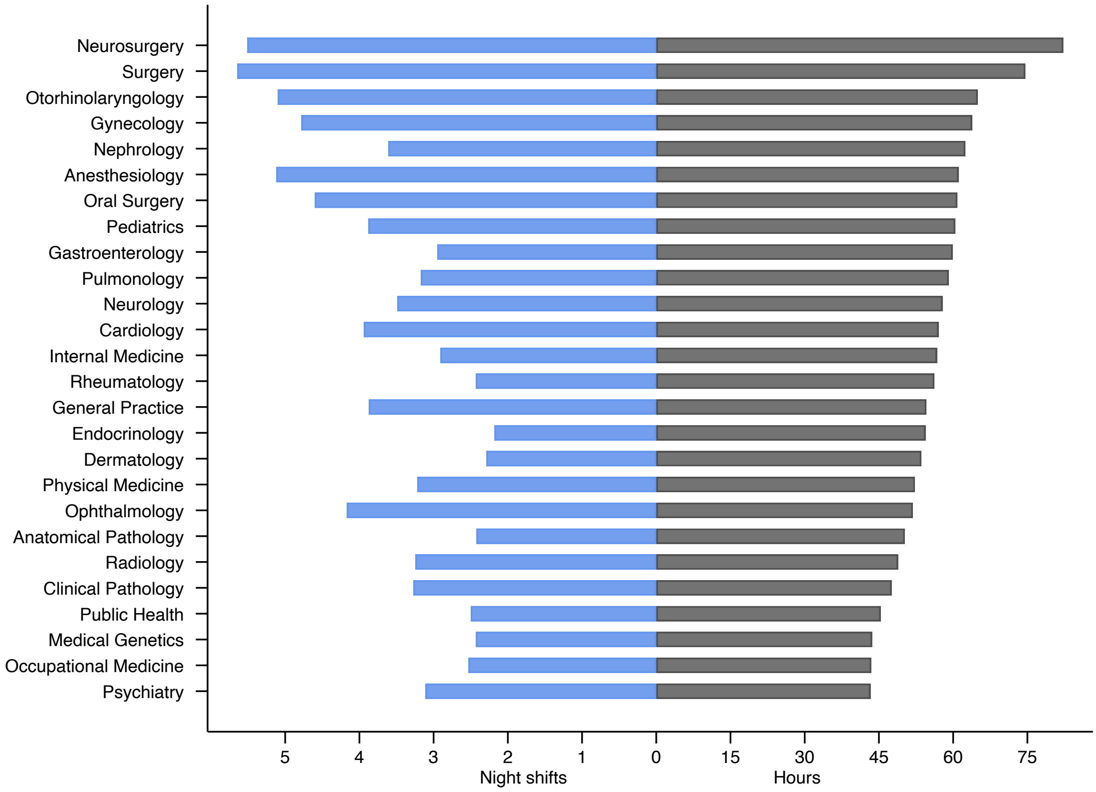
residents can work a maximum of 48 hours and participate in maximum one night shift per week.[53]

[53] Decree number 2015-225 from February 26, 2015.

To obtain variation across specialties in these two significant job attributes, we use declared hours and number of night shifts worked by medical residents. This information is obtained from two surveys, which were conducted by the largest trade union for medical residents (ISNI) on a representative sample of between 20 and 25 percent of active residents in 2019 and 2012, respectively. We link this information to the NRE data using specialty of practice.

Figure A4 plots the average number of hours worked in a week and the number of night shifts worked in a month by residents in each specialty. It highlights the variation that exists in these two dimensions across specialties: residents in neurosurgery and surgery report working almost twice as many hours as those in psychiatry and occupational medicine. Similarly, residents in surgery do more than 5 night shifts a month, while those in endocrinology do 2.

Figure A4: Night Shifts and Hours Worked by Specialty.



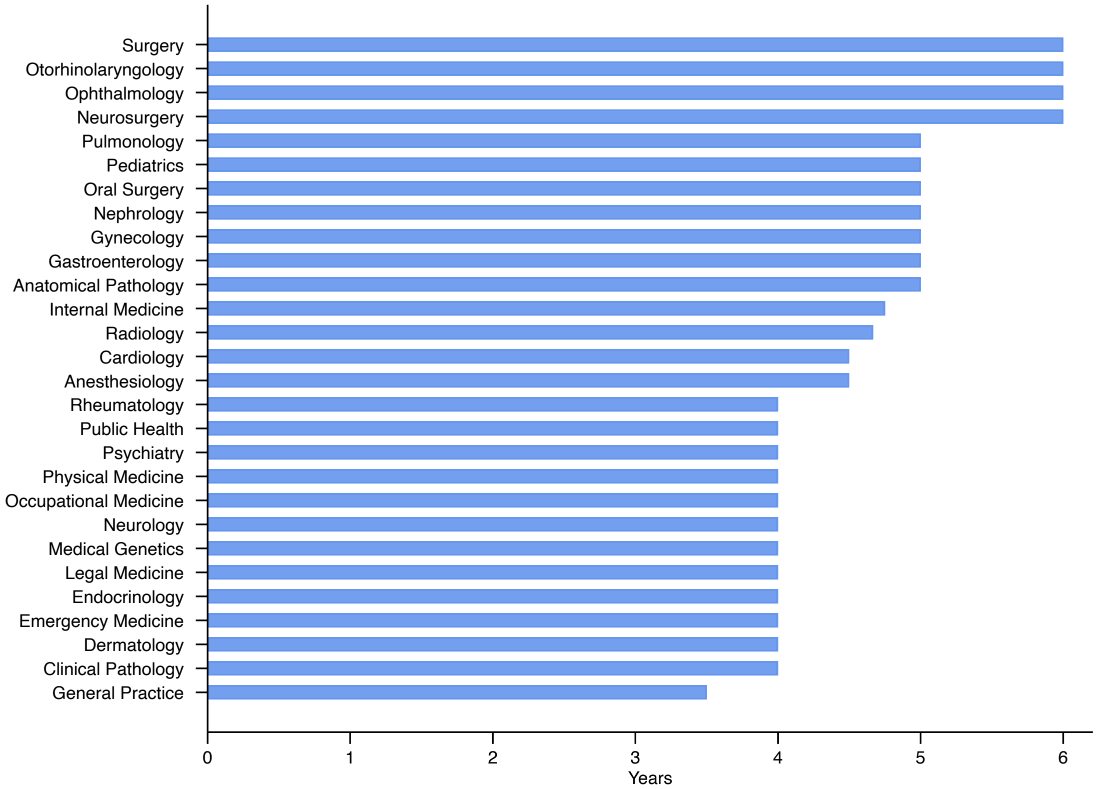
Notes: This figure shows the average number of hours worked in a week by medical residents in each specialty (left panel) and the average number of night shifts performed in a month by medical residents in each specialty (right panel). The data comes from two different surveys conducted to a nationally representative sample of medical residents in 2019 and 2012, respectively, by the trade union ISNI (*Intersyndicale nationale des internes*).

We also look at the number of years that are required for medical students to complete their residency program, as it might be an important short-term determinant of specialty choice. Figure A5 shows average residency program length by specialty. Residency length ranges from three years in general practice[54] to six years in surgery, otorhinolaryngology, ophthalmology and neurosurgery. Again, we notice some

[54]General practice is the only specialty that requires only three years of specialisation. Its average of 3.5 years results from the grouping of specialty that we use throughout the paper in which general practice and geriatrics are put together.

heterogeneity across specialties.

Figure A5: Residency Length by Specialty.



Notes: This figure shows the average number of years required to complete one's residency program by specialty. The data comes from the following 2021 ISNI report: http://www.futur-interne.com/wp-content/uploads/2021/07/ISNI-GUIDE-2021.pdf.

Other Perceived Characteristics The Occupational Information Network (O*NET) database contains hundreds of occupation-specific characteristics on a large number of occupations, including several medical specialties. Specifically, it gathers information on the knowledge, skills, and abilities required in each occupation, as well as on the activities and tasks performed, and their importance in each occupation. It also contains information on the context in which work is done, and on the values that are important in each occupation. Importantly, this data is collected via interviews to random samples of workers in each targeted occupation in the United States. One might worry that knowledge, work conditions and work values are likely to differ across countries. Even though we acknowledge that absolute differences in these characteristics across specialties are likely to differ between France and the U.S., we claim that *relative* differences should be rather stable between the two.

To link the O*NET and NRE data, we identify the relevant medical specialties in the O*NET database, and find a match for 92.5 percent of our sample.[55] Following Cortes and Pan (2018), we create four composite indices to characterise the different medical specialties, as described Table A3. First, we select

[55]Specifically, cardiology, hematology, legal medicine, nephrology, oncology, and otorhinolaryngology are not present in O*NET.

the O*NET measures which we find both relevant in our setting, and likely to be comparable in the U.S. and French contexts. We normalise each measure to have a mean of zero and a standard deviation of one in the sample of medical specialties, and then take the average of the normalised measures. We obtain four indices (competition, social component, time pressure, and interactional skills) that have mean zero and standard error of one, and which we use to characterise the occupations into which men and women self-select.

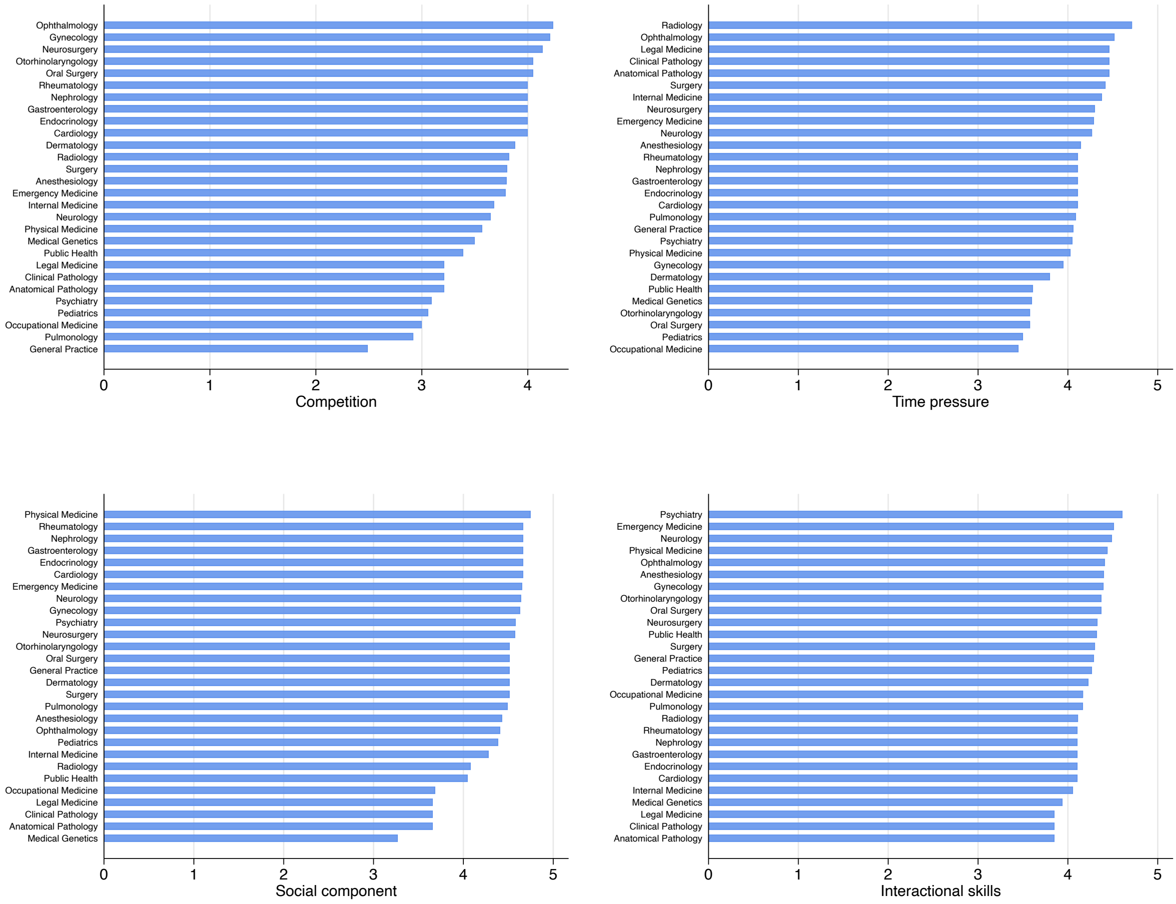
Table A3: Selected Specialty Characteristics from O*NET.

Index	Measure	Question	Scale
Time pressure	Time pressure	"How often does your current job require you to meet strict deadlines?"	From never, to every day.
Competition	Competition	"How competitive is your current job?"	From not at all, to extremely.
Social contribution	Concern for others	"How important is concern for others to the performance of your current job?"	From not at all, to extremely.
	Assisting and caring	"How important is assisting and caring for others to the performance of your current job?"	From not at all, to extremely.
	Social orientation	"How important is social orientation to the performance of your current job?"	From not at all, to extremely.
Interactional skills	Contact with others	"How much contact with others is required to perform your current job?"	From no contact, to constant contact.
	Work with a group or team	"How important are interactions that require you to work with or contribute to a work group or team to perform your current job?"	From not at all, to extremely.
	Interpersonal relationships	"How important is establishing and maintaining interpersonal relationships to the performance of your current job?"	From not at all, to extremely.
	Social perceptiveness	"How important is social perceptiveness to the performance of your current job?"	From not at all, to extremely.

Notes: The O*NET database contains hundreds of standardised and occupation-specific characteristics on almost 1,000 occupations covering the entire U.S. economy. It is updated by ongoing surveys to workers sampled from each occupation's worker population and occupation experts. From this database, we select the measures that we find relevant in this framework. All the measures are on a five-point scale. These measures are either taken as they are or aggregated into broader indices. We re-scale all our four indexes to have mean zero and a standard deviation of one, namely time pressure, competition, social contribution, and interactional skills.

Figure A6 displays the four indices by specialty, before standardisation. Again, it highlights the variation that exists across specialties.

Figure A6: Perceived Characteristics by Specialty.



Notes: This figure shows the average level of competition, time pressure, social component and interactional skills in each medical specialty as defined in O*NET.

Appendix B Supplementary Results on the Female-to-Male Earnings Gap

Table B1: Female-to-Male Earnings Gap in the Regulated-Billing Sector.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Log earnings per procedure							
Female	-0.074***	-0.088***	-0.098***	-0.084***	-0.069***	-0.039***	-0.038***	-0.040***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)
Mean dep. variable	3.484	3.484	3.484	3.484	3.484	3.484	3.484	3.484
Adj. R-squared	0.005	0.055	0.101	0.150	0.291	0.468	0.492	0.468
Observations	312,153	312,153	312,153	312,153	312,153	312,153	312,153	312,153
	Panel B: Log yearly earnings							
Female	-0.342***	-0.340***	-0.354***	-0.349***	-0.328***	-0.323***	-0.323***	-0.304***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mean dep. variable	11.763	11.763	11.763	11.763	11.763	11.763	11.763	11.763
Adj. R-squared	0.033	0.044	0.076	0.246	0.322	0.358	0.362	0.361
Observations	312,153	312,153	312,153	312,153	312,153	312,153	312,153	312,153
Year & UA size FEs	-	✓	✓	✓	✓	✓	✓	✓
Age & experience	-	-	✓	✓	✓	✓	✓	✓
Type of practice	-	-	-	✓	✓	✓	✓	✓
Procedure type	-	-	-	-	✓	-	✓	✓
Specialty FEs	-	-	-	-	-	✓	✓	✓
Family status	-	-	-	-	-	-	-	✓

Notes: Gender gap in the log of gross yearly and per procedure earnings estimated from equation (1) using all physicians with positive gross earnings billed in Sector 1. Gross yearly earnings is the sum of all the fees received for medical consultations and interventions in a year, while per-procedure earnings refers to the ratio of gross yearly earnings and the number of consultations and interventions performed in that year. "Year & UA size FEs" are fixed effects for the size of the urban area in which the practice is located, "Age & experience" includes indicators for 5 age categories as well as the number of years of experience as a private practitioner and its squared value, "Type of practice" refers to a set of dummy variables for the legal status of the practice and the share of yearly earnings arising from self-employment, "Procedure type" is the share of yearly procedures that are consultations, "Specialty FEs" refers to 16 indicators variables for each medical specialty, and "Family status" are 5 indicator variables for civil status, number of dependants and the presence of children below 5 years of age. Heteroskedasticity-robust standard errors reported in parenthesis, ***, **, * denoting significance at the 1%, 5% and 10% levels, respectively.

Table B2: Female-to-Male Earnings Gap in the Free-Billing Sector.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Log earnings per procedure							
Female	-0.243***	-0.264***	-0.284***	-0.263***	-0.232***	-0.090***	-0.080***	-0.091***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mean dep. variable	4.126	4.126	4.126	4.126	4.126	4.126	4.126	4.126
Adj. R-squared	0.034	0.092	0.180	0.198	0.238	0.566	0.580	0.567
Observations	101,247	101,247	101,247	101,247	101,247	101,247	101,247	101,247
	Panel B: Log yearly earnings							
Female	-0.383***	-0.375***	-0.413***	-0.422***	-0.362***	-0.393***	-0.376***	-0.373***
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Mean dep. variable	11.967	11.967	11.967	11.967	11.967	11.967	11.967	11.967
Adj. R-squared	0.035	0.049	0.112	0.384	0.446	0.482	0.500	0.486
Observations	101,247	101,247	101,247	101,247	101,247	101,247	101,247	101,247
Year & UA size FEs	-	✓	✓	✓	✓	✓	✓	✓
Age & experience	-	-	✓	✓	✓	✓	✓	✓
Type of practice	-	-	-	✓	✓	✓	✓	✓
Procedure type	-	-	-	-	✓	-	✓	✓
Specialty FEs	-	-	-	-	-	✓	✓	✓
Family status	-	-	-	-	-	-	-	✓

Notes: Gender gap in the log of gross yearly and per procedure earnings estimated from equation (1) using all physicians with positive gross earnings billed in Sector 2. Gross yearly earnings is the sum of all the fees received for medical consultations and interventions in a year, while per-procedure earnings refers to the ratio of gross yearly earnings and the number of consultations and interventions performed in that year. "Year & UA size FEs" are fixed effects for the size of the urban area in which the practice is located, "Age & experience" includes indicators for 5 age categories as well as the number of years of experience as a private practitioner and its squared value, "Type of practice" refers to a set of dummy variables for the legal status of the practice and the share of yearly earnings arising from self-employment, "Procedure type" is the share of yearly procedures that are consultations, "Specialty FEs" refers to 16 indicators variables for each medical specialty, and "Family status" are 5 indicator variables for civil status, number of dependants and the presence of children below 5 years of age. Heteroskedasticity-robust standard errors reported in parenthesis, ***, **, * denoting significance at the 1%, 5% and 10% levels, respectively.

Appendix C Supplementary Results on the NRE

C.1 Gender Gaps in Occupational Choices on the Full Sample

In addition to estimating the gender gap in occupational sorting among top performers reported in Section 4.2 Figure 2, we also estimate the gender gap in the probability of self-selecting into each speciality for the whole support of the exam score distribution. We report the results in Figure C1, in which three different patterns emerge. First, some specialties exhibit no gender differences in self-selection, regardless of the position in the rank distribution. Second, a group of specialties feature a gender gap that increases—in absolute terms—with exam score. Specialties such as cardiology, dermatology and ophthalmology thus have their largest gender gap in the upper part of the exam score distribution.

Third, gender gap estimates and exam score have a U-shaped relationship in anesthesiology, radiology, and surgery, and an inverse U-shaped relationship in general practice, gynecology, and pediatrics. It suggests that, in the latter group of specialties, the gender gap in self-selection is larger in absolute terms towards the middle of the exam score distribution than at the top and bottom. The two extreme examples are general practice and surgery. While women in the top 5 percent of the performance distribution are 5 percentage points more likely to choose general practice than the men in the same percentiles, women between the 50th and the 60th percentile are 12 percentage points more likely to choose general practice than their male counterparts who face the same choice set. Symmetrically, while women in the top 5 percent of the exam score distribution are 5 percentage points less likely to choose surgery than men, women between the 50th and the 60th percentile are 12 percentage points less likely to choose surgery than their male counterparts facing the same choice sets.

This suggests that, for some specialties, as we move our focus towards lower parts of the exam score distribution where choice sets are substantially smaller, gender differences in occupational choices are exacerbated. One potential explanation for this pattern is the existence of gender differences in students' willingness to choose their preferred specialty when it comes at the cost of having to choose a less preferred location for their medical residency. Indeed, as all the vacancies for one's preferred specialty are filled in one's preferred location, an occupation-location trade-off emerges. We test this hypothesis in our survey by asking students what their choice would be if they were faced with such a trade-off during their up-coming residency choice, and we do not find statistically significant differences in their declared choices. It suggests that students who populate the lower parts of the exam score distribution have different preferences for specialties than those at the top. As a matter of fact, exam score relates to relevant individual characteristics. For instance, for both genders, exam score is negatively related to the probability of being married and positively correlated with parental education.

Figure C1: Gender Gap in Self-selection into Specialties Across the Exam Rank Distribution.

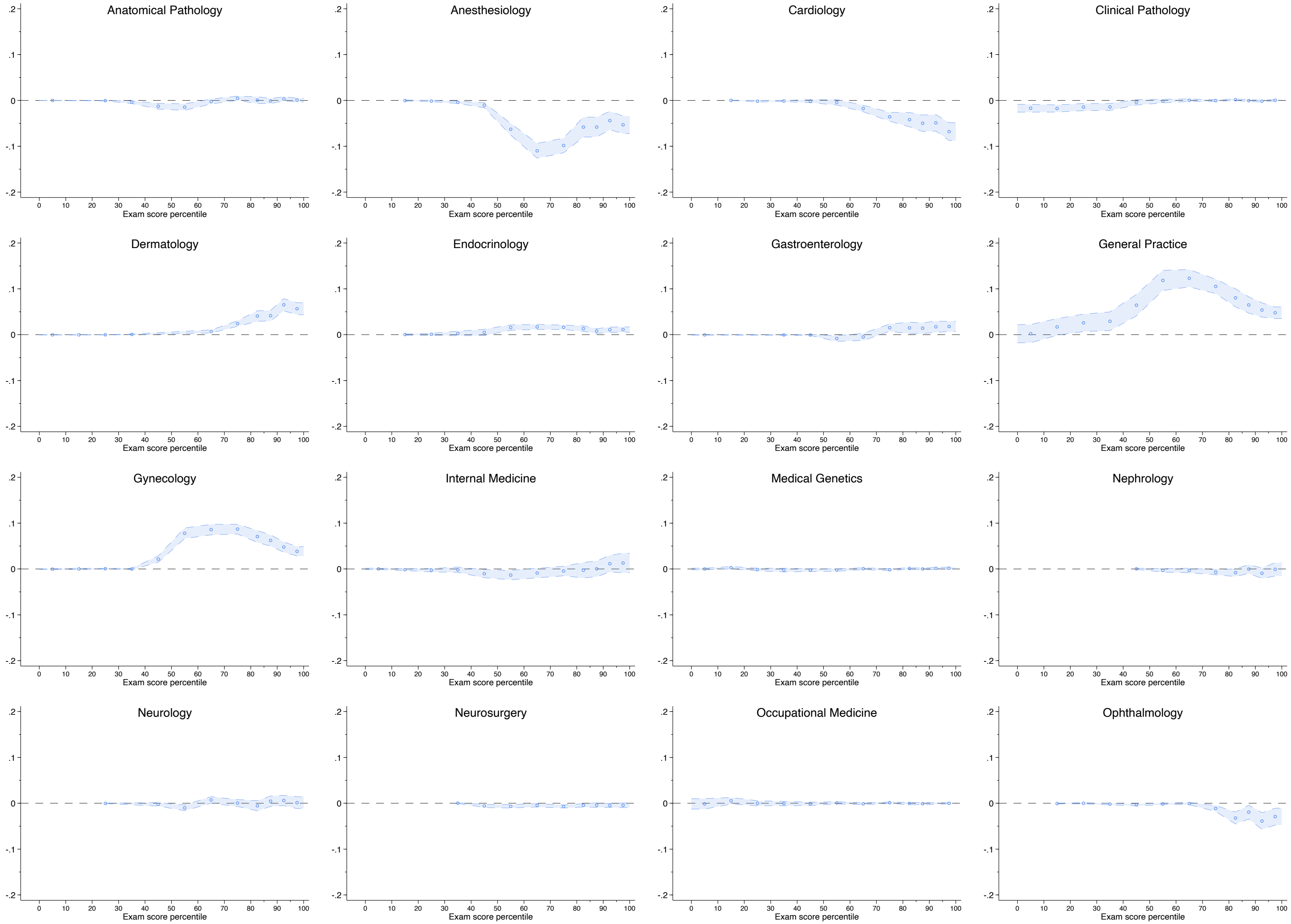
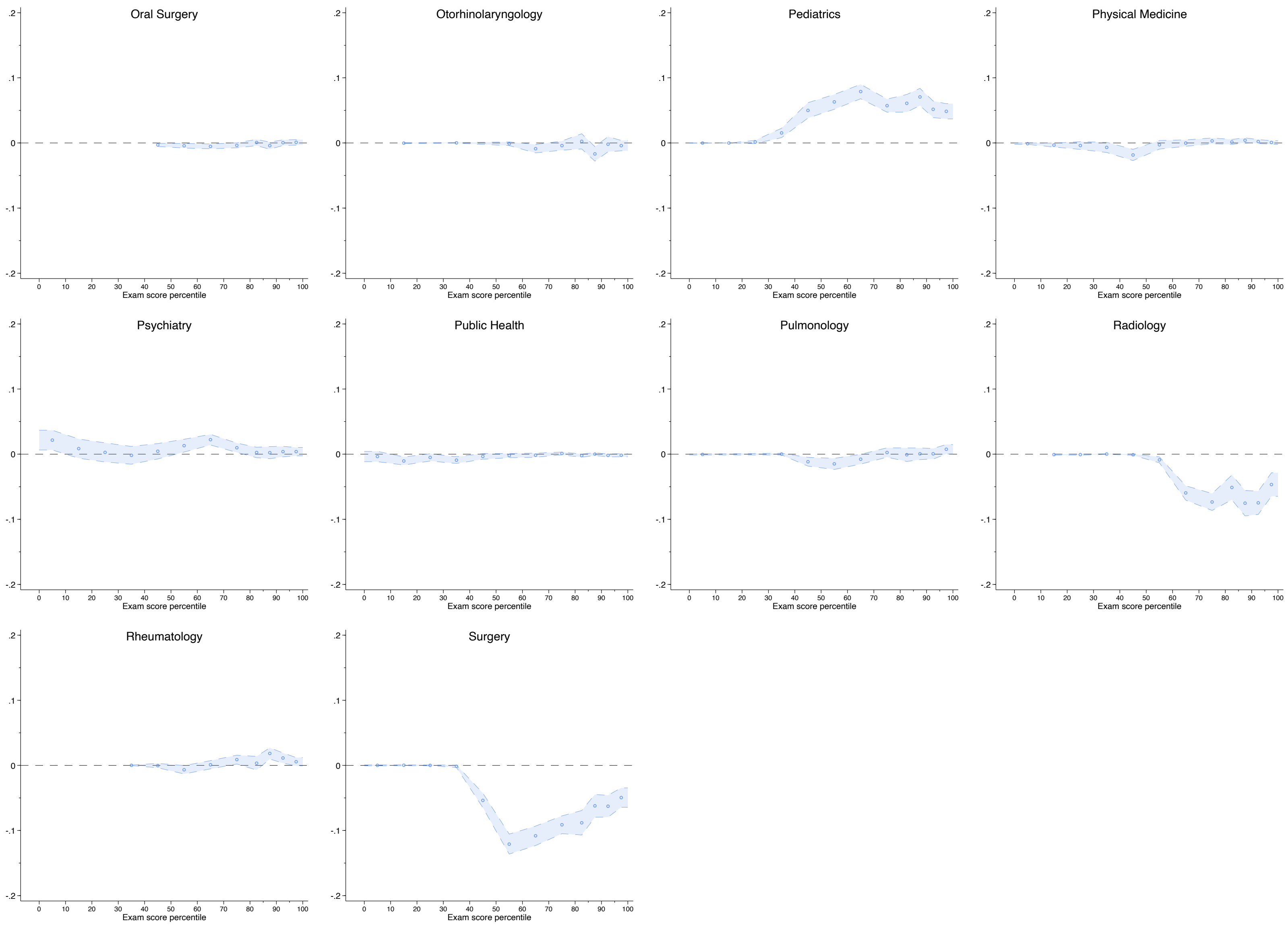


Figure C1 (cont.): Gender Gap in Self-selection into Specialties Across the Exam Rank Distribution.



Notes: This figure plots the coefficients $\hat{\beta}$ and 95% confidence intervals in equation (2) estimated separately in each bottom 8 decile and top 4 ventile of the exam score distribution, for 26 medical specialties. Negative (positive) values indicate that women are less (more) likely to self-select into the corresponding specialty than men with a virtually identical choice set, on average. Missing estimated gender gaps in some parts of the distribution are due to no student choosing that specialty. All the gender gaps are estimated using OLS and include choice set fixed effects defined as groups of 5 individuals with consecutive exam scores. The shaded area shows 95% confidence intervals using heteroskedastic robust standard errors.

C.2 Robustness Checks

In this section, we perform a series of sensitivity tests to assess the robustness of our results. First, we show that our baseline results are robust to an alternative definition of groups within which choice sets are considered as identical. Second, we show that the result according to which men and women who are unconstrained by position availability differ in their occupational choices is robust to alternative definitions of unconstrainedness. Third, we show that the results go through when using a modified regression framework that accounts for each specialty's relative (female) popularity.

C.2.1 Alternative Definition of Groups with Similar Choice Sets

So far, we have considered the choice sets of five individuals with consecutive ranks as similar. Within such groups, the choice set of the best ranked and the worst ranked can at most differ by 4 positions. We now use two alternatives, more restrictive definitions of these groups: (i) groups of individual facing exactly the same choice sets, and (ii) pairs composed of a man and a woman of consecutive rank. Using these definitions, we re-run the baseline analysis described in equation (2). The results are shown in Figure C2, and comparing them to those displayed in Figure 2 shows that our definition of choice set fixed effects is not driving our results.

C.2.2 Alternative Definitions of Being Unconstrained by Position Availability

Being unconstrained by position availability is the fact of making the occupational decision while facing no formal barriers to entry into one's preferred position. As a result, unconstrained jobseekers are those who pick their job while 100 percent of positions are still available. Given that this definition is restrictive in terms of number of observations and thus largely decreases the statistical power of our analysis, we have instead decided throughout the paper to use as the unconstrained sample the group of top 5 percent performers at the National Ranking Examinations each year. In this section, we show that this choice does not drive our results, by replicating Figure 2 using the following alternative definitions: (i) making a decision while 100 percent of the positions are still available (*strict* unconstrainedness); (ii) making a decision while 99.5 percent of the positions are still available; (iii) making a decision while 99 percent of the positions are still available. The results are displayed in Figure C3. Even though the results are qualitatively slightly different, the main result according to which men and women who are unconstrained differ in their occupational choices is largely unchanged.

C.2.3 Alternative Regression Framework

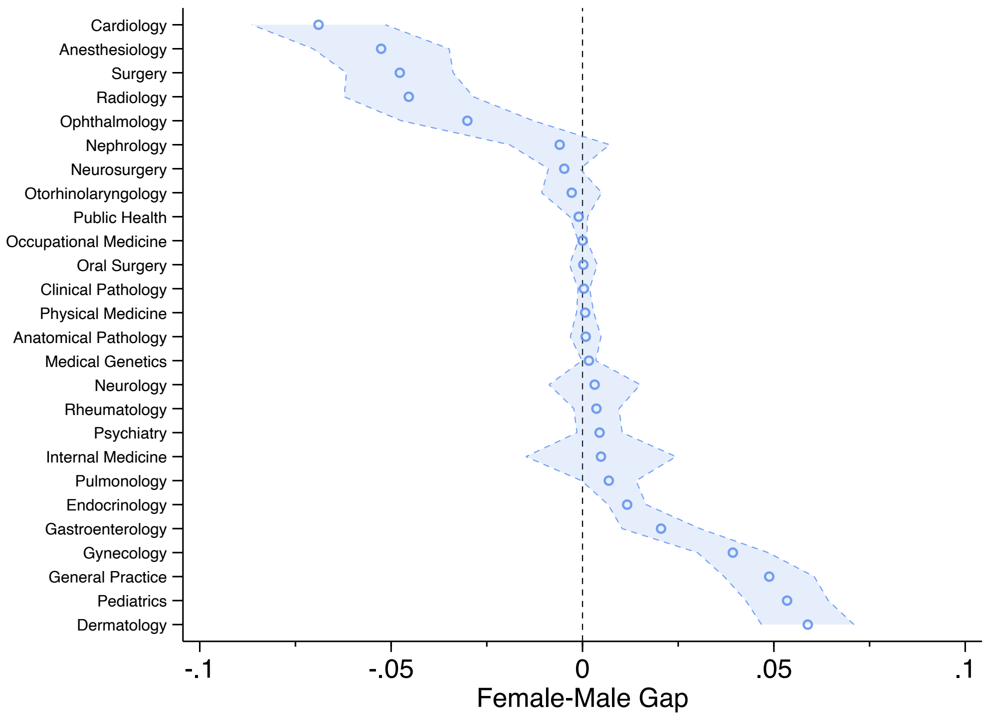
Under our regression framework, the magnitude of each specialty's estimated gender gap is bounded by its relative popularity.[56] To understand how the estimated gaps change when accounting for each specialty's relative (female) popularity, Figure C4 plots the point estimates from equation (2) expressed as the proportion of the share of women who choose each specialty. The relative popularity of each specialty in the population of top 5 percent exam performers is also shown by its marker's size. Even though the magnitude and interpretation of the gender gaps change with this transformation, the ordering of the specialties as well as the conclusions remain very similar. Indeed, the specialties that see their gender gap increase the most (in absolute terms) are among the least popular, and their gender gaps remain statistically insignificant.[57]

[56]Taking a fictitious specialty that is only selected by 1 percent of the students, this means that the estimated gender gap cannot be larger than 1 percent. We are grateful to anonymous referee 1 for pointing this out and for their advice on an alternative measure.

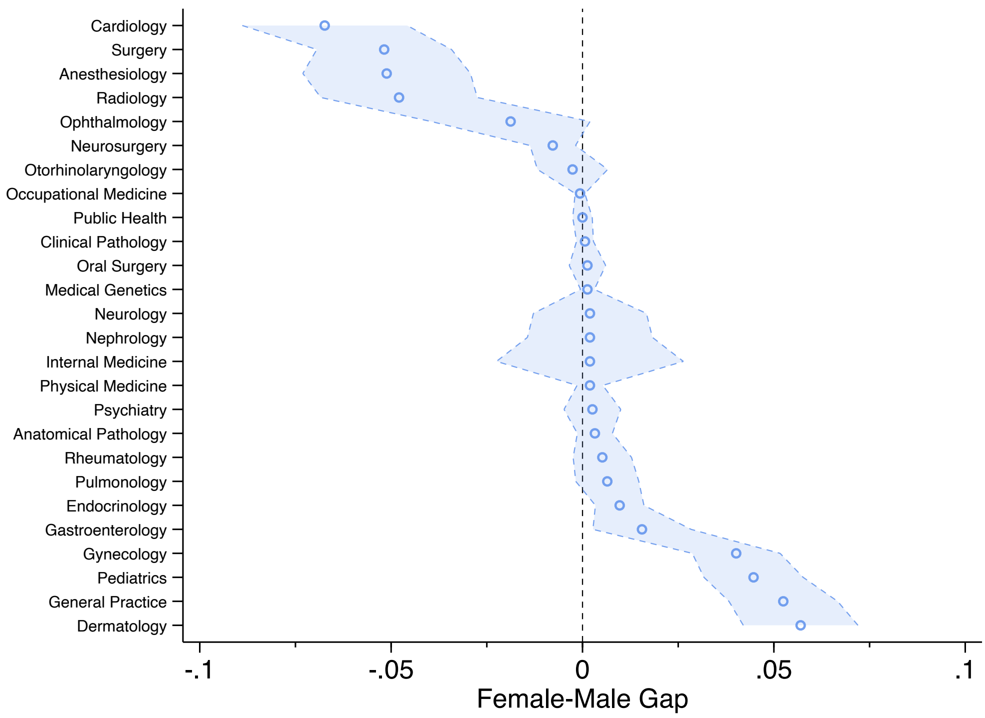
[57]The one exception is endocrinology, which exhibit a positive and significant gender gap in Figure 2 as well.

Figure C2: Gender Differences in Self-selection into Specialties, Using Alternative Definitions of Choice Set Fixed Effects.

(a) Groups with strictly identical choice sets.



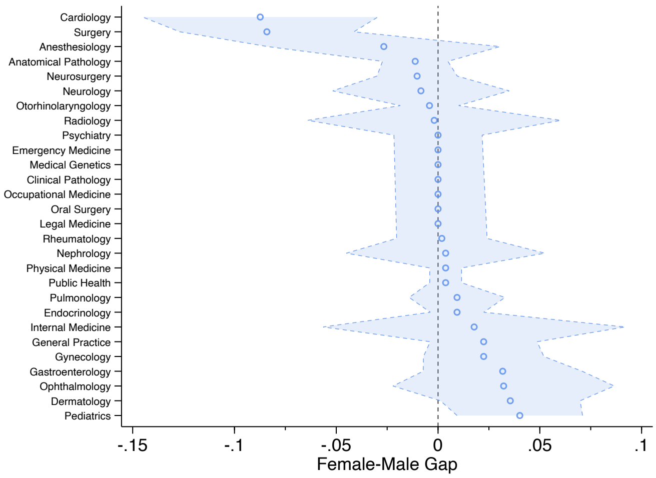
(b) Pairs of one man and one woman with consecutive exam rank.



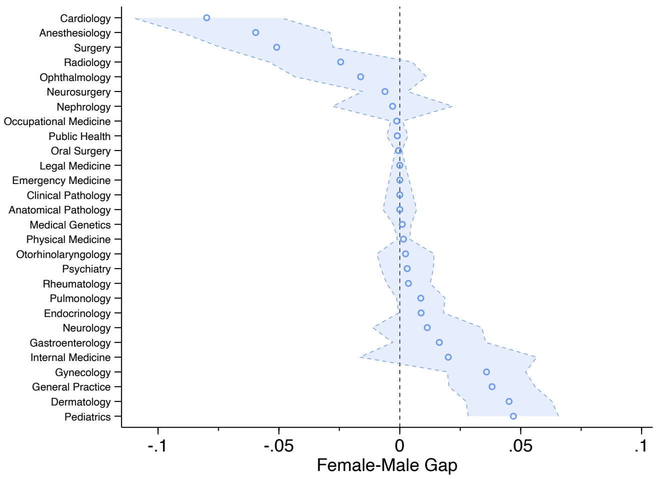
Notes: This figure plots the OLS estimates for the coefficients $\hat{\beta}$ and 95% confidence intervals estimated from equation (2) separately by specialty, using (a) groups with strictly identical choice sets, and (b) pairs of one man and one woman of consecutive exam scores as alternative definitions for choice set fixed effects. The shaded area shows 95% confidence intervals using heteroskedastic robust standard errors.

Figure C3: Gender Differences in Self-selection into Specialties, Using Alternative Definitions of Unconstrainedness.

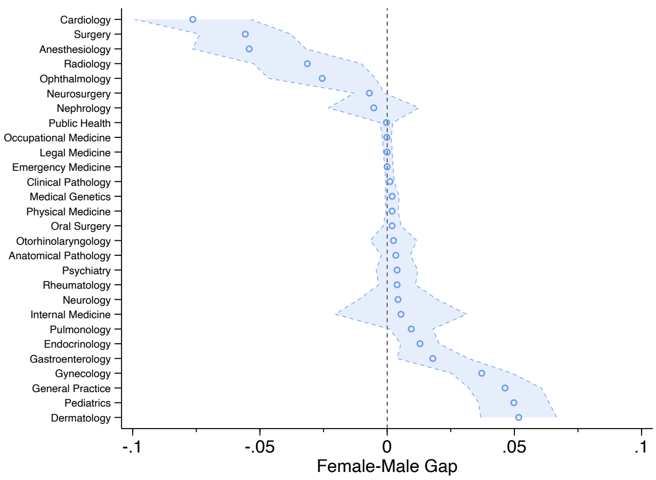
(a) 100% of positions are available.



(b) 99.5% of positions are available.

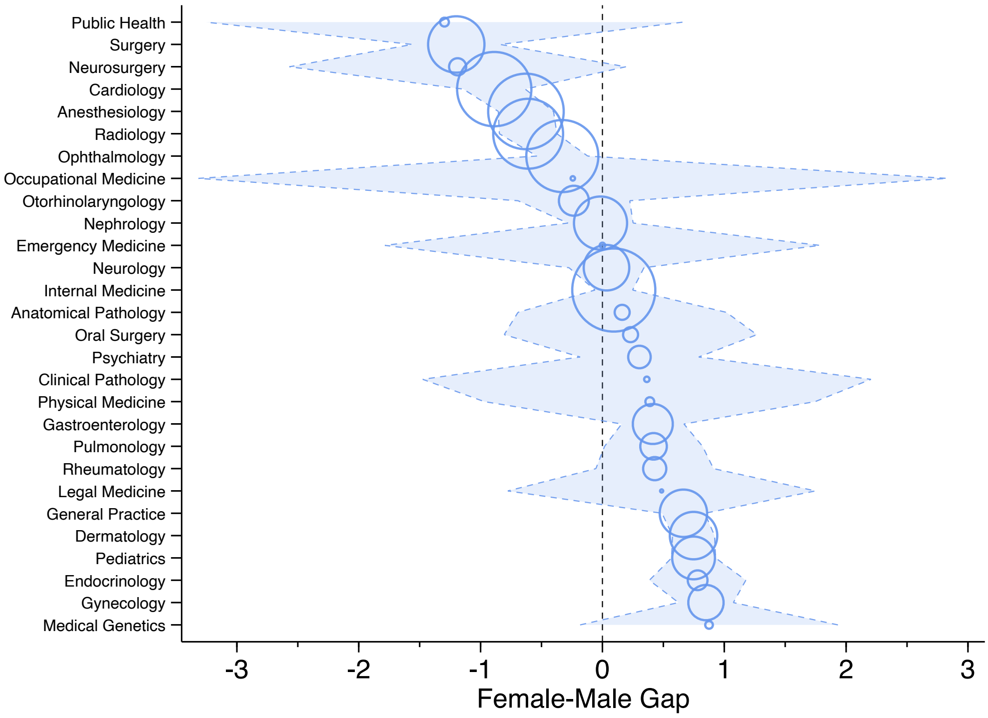


(c) 99% of positions are available.



Notes: This figure plots the OLS estimates for the coefficients $\hat{\beta}$ and 95% confidence intervals estimated from equation (2) separately by specialty, focusing on (a) candidates who make their choice when all positions are still available, (b) candidates who make their choice when 99.5 percent of positions are still available, and (c)candidates who make their choice when 99 percent of positions are still available. All regressions are estimated by OLS and include choice set fixed effects defined as groups of 5 individuals with consecutive exam scores. The shaded area shows 95% confidence intervals using heteroskedastic robust standard errors.

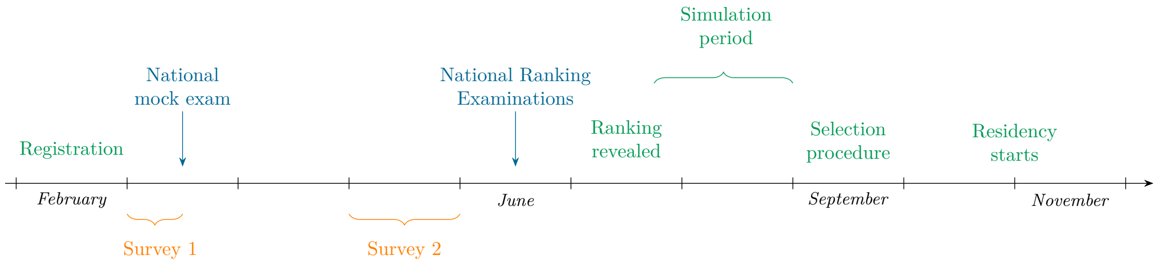
Figure C4: Adjusted Gender Differences in Specialty Choices (Top 5% Exam Performers).



Notes: This figure plots the coefficients $\hat{\beta}$ and 95% confidence intervals as a proportion of the share of females in each specialty, resulting from estimating (2) for each medical specialty, using the sample of students who scored in the top 5% of the National Ranking Examinations in the years 2010-2022 ($5,078$ individuals). The size of the marker represents the relative popularity of each specialty in the population of top 5 percent exam performers. This group of students chose their residency position when 99.7% or more of all positions were still available. Negative (positive) values indicate that, on average, women are less (more) likely to self-select into the corresponding specialty than men with a virtually identical choice set. All the gender gaps are estimated using OLS and include choice set fixed effects defined as groups of 5 individuals with consecutive exam scores. The shaded area show 95% confidence intervals using heteroskedastic robust standard errors.

Appendix D Supplementary Descriptive Material on the Survey

Figure D1: Project Timeline.



Notes: This Figure provides an overview of our project's timeline. The dates that relate to the organisation of the National Ranking Examinations are shown in blue and green. Note that they are based on the 2022 NRE, and that they might change slightly in different years. The dates that relate to our data collection are shown in orange.

Table D1: Follow-up Survey Respondents' Demographic Characteristics.

	(1) Women	(2) Men	(3) Diff.	(4) *p*-value	(5) Obs.
Age	24.32	24.64	-0.32***	0.005	1,381
French	0.99	0.96	0.03***	0.000	1,382
With children	0.01	0.02	-0.01	0.195	1,382
Marital status:					
Single	0.39	0.39	0.01	0.805	1,236
In a relationship	0.56	0.58	-0.03	0.403	1,236
Married or in a civil partnership	0.05	0.03	0.02	0.197	1,236
Divorced	0.00	0.00	0.00	0.328	1,236
Fertility plans for the next 5 years:					
Wants children in the next 5 years	0.30	0.21	0.09***	0.001	1,297
Does not want children in the next 5 years	0.35	0.42	-0.07	0.018	1,297
Does not know	0.35	0.37	-0.02	0.419	1,297
Type of baccalauréat:					
Scientific	0.99	0.98	0.00	0.928	1,382
Non-scientific	0.01	0.02	-0.00	0.928	1,382
Baccalauréat honors:					
Highest honors or Very high honors	0.62	0.49	0.14***	0.000	1,379
High honors or Honors	0.36	0.48	-0.12***	0.000	1,379
No honors	0.02	0.04	-0.02**	0.032	1,379
Mother's education:					
High-school diploma or below	0.27	0.34	-0.07**	0.011	1,322
Bachelor's or equivalent	0.29	0.27	0.02	0.384	1,322
Master's or equivalent	0.32	0.26	0.06**	0.031	1,322
PhD or equivalent	0.11	0.12	-0.01	0.485	1,322
Father's education:					
High-school diploma or below	0.32	0.37	-0.05*	0.079	1,302
Bachelor's or equivalent	0.19	0.14	0.05**	0.016	1,302
Master's or equivalent	0.35	0.31	0.04	0.146	1,302
PhD or equivalent	0.14	0.18	-0.05**	0.035	1,302
Passed the first year of medical school:					
First attempt	0.49	0.46	0.03	0.275	1,338
Second attempt	0.50	0.51	-0.01	0.695	1,338
Other	0.01	0.03	-0.02**	0.010	1,338
Rank first year of medical school:					
Top 10%	0.20	0.22	-0.02	0.512	1,301
Top 10 to 25%	0.15	0.18	-0.03	0.195	1,301
Bottom 10%	0.13	0.12	0.01	0.488	1,301

Notes: This table reports some socio-demographics characteristics on the sample of respondents of the second survey. Reported *p*-values are from a two-sample t-test of differences in means, where *** p<0.01, ** p<0.05, * p<0.1.

Table D2: Geographic Coverage of Respondents.

Region	(1) Women	(2) Men	(3) Diff.	(4) p-value	(5) Obs.
Auvergne-Rhône-Alpes	0.14	0.12	0.02	0.367	1,382
Bourgogne-Franche-Comté	0.03	0.03	-0.00	0.638	1,382
Bretagne	0.05	0.05	-0.00	0.871	1,382
Centre-Val de Loire	0.02	0.03	-0.01	0.545	1,382
Grand Est	0.10	0.12	-0.02	0.210	1,382
Hauts-de-France	0.14	0.13	0.01	0.543	1,382
Normandie	0.05	0.07	-0.01	0.277	1,382
Nouvelle-Aquitaine	0.10	0.11	-0.01	0.532	1,382
Occitanie	0.10	0.07	0.03*	0.064	1,382
Pays de la Loire	0.08	0.08	0.00	0.993	1,382
Provence-Alpes-Côte d'Azur	0.02	0.03	-0.01	0.125	1,382
Île-de-France	0.19	0.17	0.01	0.583	1,382

Notes: This table reports the share of survey respondents in the region where their university is located. Reported p-values are from a two-sample t-test of differences in means, where *** p<0.01, ** p<0.05, * p<0.1.

D.1 Expected and Realised Exam Performance

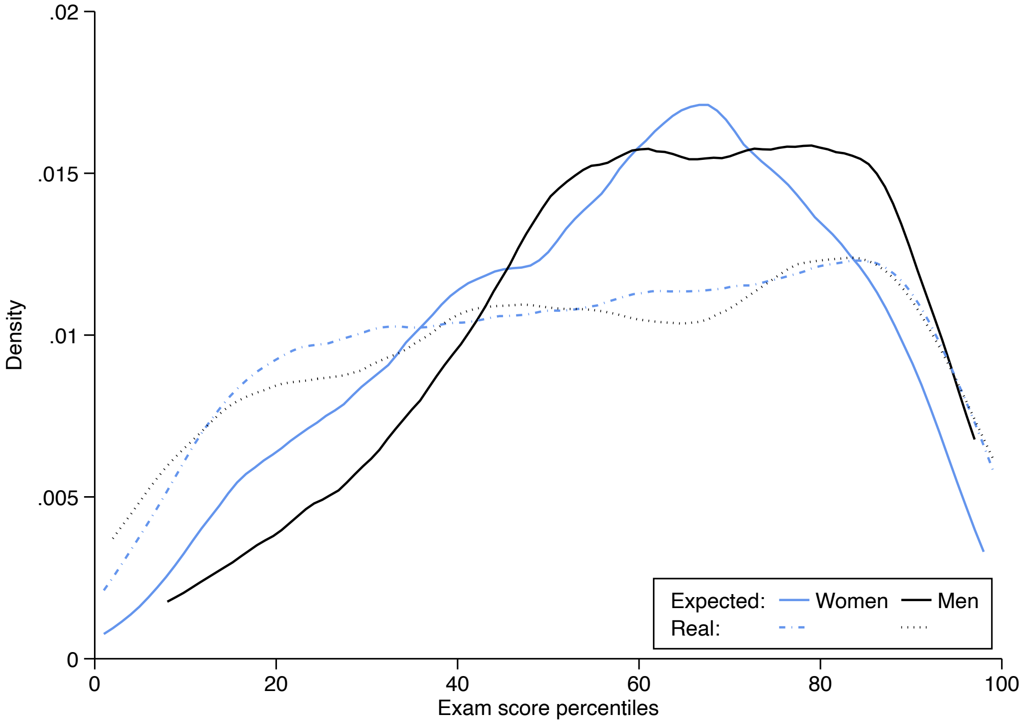
Figure D2 shows the respondents' expected and realised performances at the NRE. The latter comes from the official ranking released following the 2022 NRE.[58] The former was elicited during the survey by asking respondents to introduce the probability that their realised rank at the 2022 NRE falls into each of a series of pre-defined intervals.[59] Since there were more than 9,000 students registered to take the 2022 NRE, we provided survey respondents with 19 intervals each comprising 500 adjacent ranks. We obtained each respondent's expected exam score by adding up each interval's median value weighted by their reported marginal probability.

Figure D2 presents the density distributions of the expected and realised exam ranks, separately for each gender. It shows that, for both genders, the expected score density distribution is less skewed than that of the realised scores. While female respondents seem to be more cautious than male respondents when predicting their score, as they end up in the upper part of the distribution more often than what they expect, the difference for male respondents is smaller. In both cases, the distribution of the expected score is more skewed to the left than that of the real score.

[58] JORF n°0168, text n°52, July 22nd 2022.

[59] The usage of subjective marginal probabilities creates the usual trade-off in survey design between obtaining more detailed answers at the expense of requiring the respondents to exert a higher effort to provide them.

Figure D2: Expected and Realised Exam Score Kernel Densities by Gender.



Notes: This figure plots the estimated probability density function of the expected and realised exam score by kernel density estimation for female and male respondents who took the NRE in 2022.

Appendix E Supplementary Material on the Hypothetical Job Choice Experiment

E.1 A Model for Hypothetical Job Choices

Let i denote an individual asked to respond to a choice scenario presenting J jobs, each characterised by K attributes. Under the assumptions of linear preferences over the job attributes and of utility-maximising decision makers, we set up the canonical random-utility model in which medical student i derives utility from choosing job j in the following way:

$$U_{ij} = \beta_i X_j + \epsilon_{ij} \tag{7}$$

where X_j is a vector containing the observed attributes of job j, β_i is a vector containing individual i's utility weights over those attributes, and ϵ_{ij} is an idiosyncratic error term containing the remaining characteristics that affect respondent i's utility (it is observed by the respondent but not by the researcher).

Further assuming that ϵ_i are independent and identically distributed (i.i.d.) Type I extreme value conditional on the attributes X_j, the probability of choosing job j conditional on $X_1, ..., X_J$ is given by:

$$p_{ij} = \frac{exp(X_j'\beta_i)}{\sum_{j'=1}^{J} exp(X_{j'}'\beta_i)}, \quad j = 1, ..., J \tag{8}$$

This yieds the mixed multinomial logit of Mcfadden and Train (2000).

Note that so far, we have assumed that respondents know the attributes X_j for all j's as well as their own $\epsilon_i = \epsilon_{i1}, \epsilon_{i2}, ..., \epsilon_{iJ}$. However, given that we only provide them with information on the job attributes, it might not be realistic to assume that they know ϵ_i. By collecting choice probabilities instead of stated choices, we enable respondents to express uncertainty about ϵ_i and thus about the choice that they would make in a real choice situation. This is the "elicited-choice-probabilities" approach described by Blass, Lach and Manski (2010) and motivated in Section E.2 of this paper.

We then obtain a linear mixed-logit model by applying the log-odds transformation to equation (8):

$$\ln\left(\frac{p_{ij}}{p_{i1}}\right) = (X_j - X_1)'\beta_i, \quad j = 2, ..., J \tag{9}$$

where β_i is a vector of length K containing the marginal changes in the log odds of choosing job j over job 1 following a change in the relative levels of job j and job 1 in terms of the K attributes.[60]

As acknowledged by Blass, Lach and Manski (2010) and in turn Wiswall and Zafar (2018), one issue

[60] The alternative $j = 1$ is arbitrarily chosen.

that might plague the preference parameters estimated from hypothetical data is that individuals might report their preferences with error. For instance, individuals tend to round their reported probabilities to units of 5 and 10. This only has minor implications with interior probabilities, but rounding close to 0 and 100 generates log odd ratios that equal minus and plus infinity, and thus cause least squares estimations to break down. A way to resolve this issue is to use the linear median regression model:

$$M\left[\ln\left(\frac{p_{ij}}{p_{i1}}\right) \mid X\right] = (X_j - X_1)'\beta_i \tag{10}$$

where $M[\cdot]$ is the median operator. The advantage of this model is that, by assuming that preferences are symmetrically distributed around their median value, it only requires that measurement errors are median unbiased. It is also invariant to transformations of the extreme values (0 and 100) that do not alter the ordering of preference values around the median.

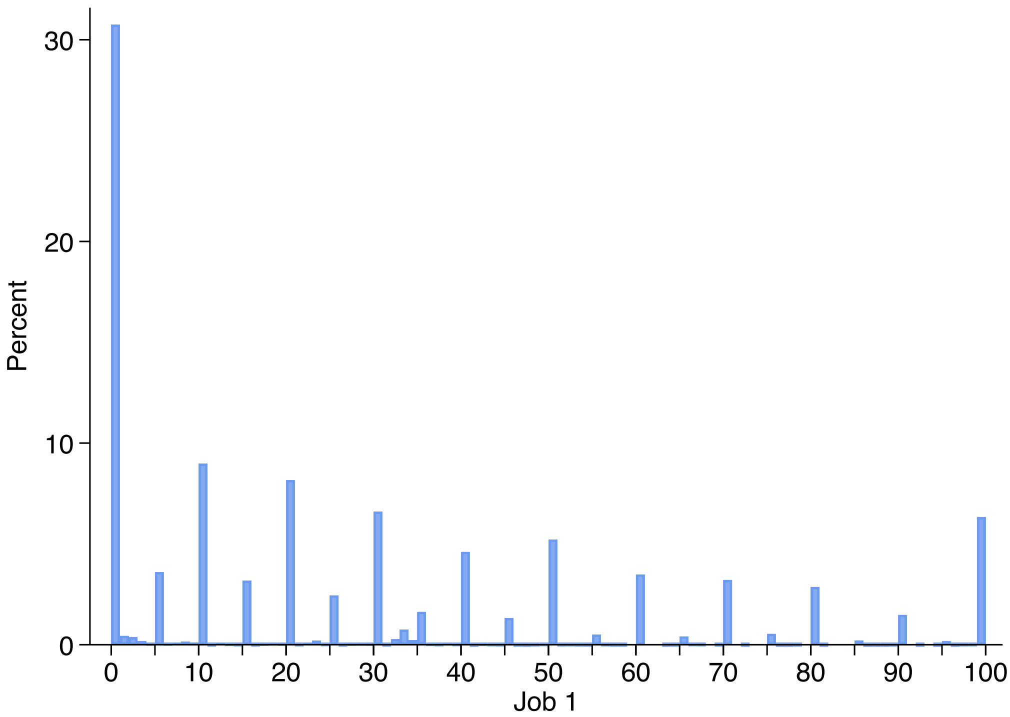
E.2 The Elicited Choice Probabilities

Figure E1 shows the distribution of the elicited choice probabilities for job 1 before sample selection. The first observation arising from this figure is that most responses are multiples of 5 and 10, which suggests that individuals tend to round their answers. Second, and more worryingly, we see that 0 is by far the most common answer, with 30 percent of respondents assigning a probability 0 to job 1.[61] Blass, Lach and Manski (2010) note that excessive rounding could be problematic. One specificity of our setting is that we survey respondents only three months before they make their actual occupational choice. In this setting, respondents are likely to have much clearer ideas about the alternatives they face during the hypothetical job choice experiment than respondents in previous studies such as Wiswall and Zafar (2018) and Blass, Lach and Manski (2010). As a result, ex-ante, we expect much more bunching at the extreme values of 0 and 100 than in previous studies. Still, one might argue that these observed extreme probabilities are also in part due to respondents who do not put enough attention and introduce values as quickly as possible. In order to minimise the extent of this problem and make sure that this bunching is due to actual preferences, we remove from our sample the individuals who spend less than 3 minutes on the job choice experiment. Figure E2 shows that these individuals are much more likely to input extreme probabilities than the rest of the respondents, which provides evidence of their inattention.

The third observation to draw from Figure E1 is that, even though 0 is the most frequent answer, a large share of the elicited probabilities are interior probabilities. Specifically, focusing on job 1, more than 60 percent of the elicited choice probabilities are interior, which suggests that individuals do feel the need to express uncertainty about their choice.

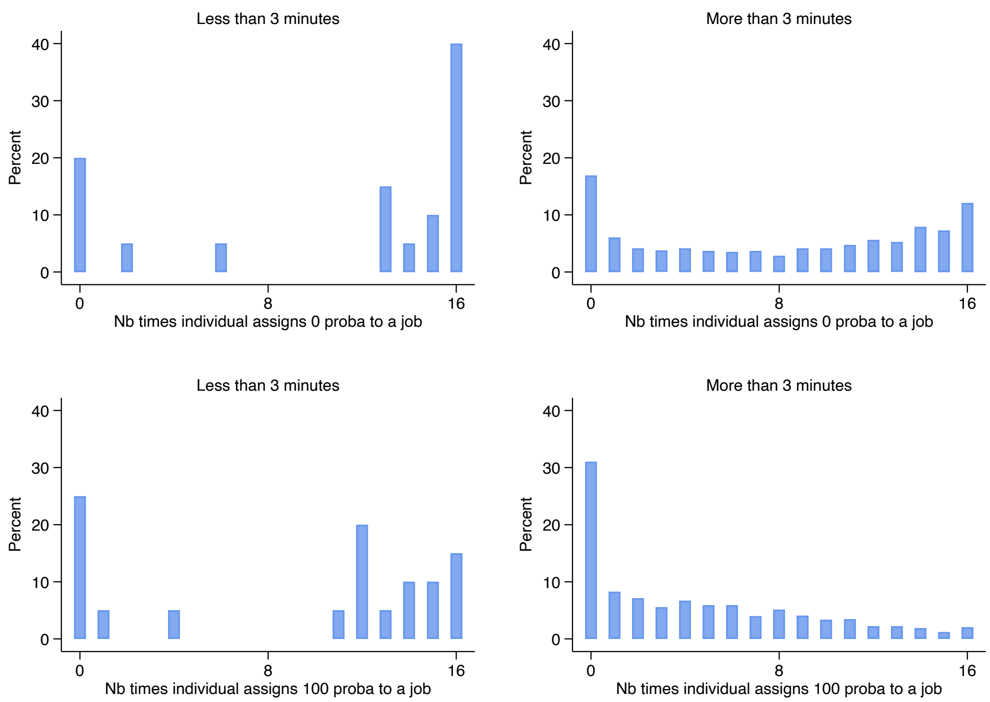
[61] The picture looks similar for jobs 2 and 3. 25 and 23 percent of respondents input a 0 probability to job 2 and 3, respectively.

Figure E1: Elicited Choice Probabilities for Job 1.



Notes: This figure shows the distribution of choice probabilities for job 1 elicited during the hypothetical job choice experiment.

Figure E2: Number of Respondents who Assign an Extreme Probability to a Job.



Notes: This figure shows the number of times respondents assign an extreme probability to a job, depending on how much time they spent on the hypothetical job choice experiment. The figures on the top row focus on 0 while the figures on the bottom row focus on 100. Given that there are 16 scenarios in total, each individual can input a given probability at most 16 times.

E.3 Supplementary Hypothetical Job Choice Experiment Results

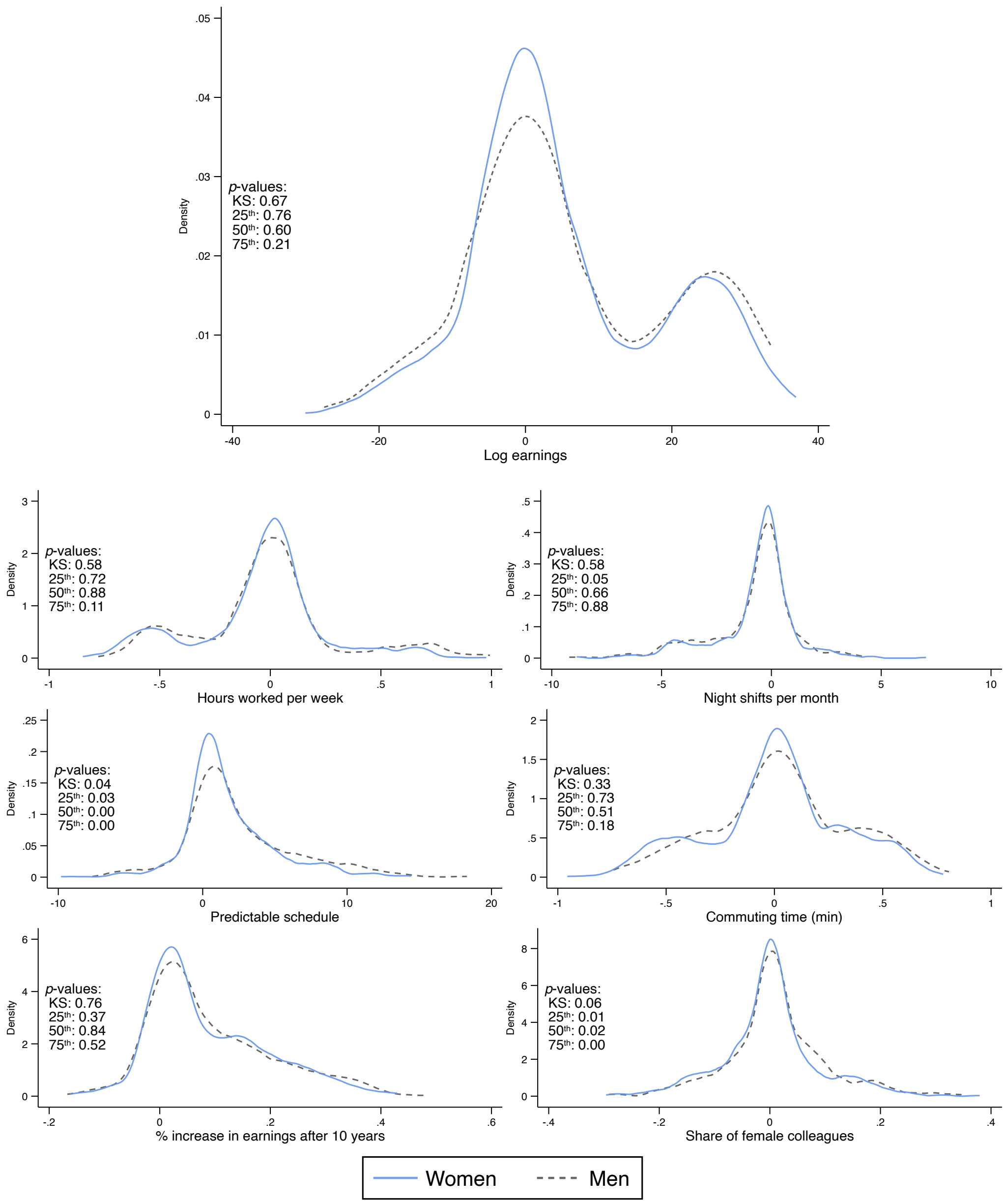
Figure E3 displays the estimated kernel densities of the job preference parameters β from job choice model (4) by gender. We complement these figures with some descriptive statistics on the estimates in Table E1. It first shows that, unsurprisingly, most respondents prefer jobs with higher earnings, as the distributions' median values are positive for both genders, and that a significant share of respondents have a strong taste for higher earnings (bunching around 25 log points). We also find that a non-negligible number of respondents have zero or negative preference estimates for earnings.[62] Although we fail to reject the null hypothesis of equality of the two distributions through the Kolmogorov-Smirnov test, testing for gender differences at different parts of the distributions sheds light on the existence of significant gender differences. It reveals that men are more concentrated than women on the right side of the distribution, suggesting that men more often have a strong taste for higher earnings than women.

E.3.1 Descriptive Statistics

The mean and median values of the estimated preferences for the other job attributes have the expected signs (Table E1). Most respondents dislike working more hours and night shifts, and having to do longer commutes, while they show a taste for having a predictable schedule and higher earnings growth after 10 years. It also appears that most respondents are indifferent to the share of female colleagues in their workplace, as shown by the high concentration of estimated preferences around 0. Keeping in mind that interpreting the magnitude of these estimates is not straightforward, we turn to analysing the amount that each individual is willing to pay for each of these job attributes.

[62] Given the specificity of the medical setting, we did expect a significant number of respondents to be indifferent to earnings. The number of individuals with negative preference for earnings is however somewhat of a surprise, and this is why we exclude the bottom 5 percent of the estimated earnings preferences distribution in the main analysis. Note that this restriction does not significantly affect the results.

Figure E3: Kernel Densities of Estimated Preferences for Job Attributes.



Notes: This figure displays the kernel densities of each utility weight β estimated from equation (4) by gender. KS denotes p-values for Kolmogorov-Smirnov tests of the equality of the two distributions. We also conduct tests for gender differences in the 25^{th}, 50^{th} and 75^{th} percentiles using bootstrapped samples.

Table E1: Descriptive Statistics on the Estimates from the Job Choice Model.

		Overall	Men	Women
Log earnings				
	Mean	5.80***	6.22***	5.61***
	25th pct.	-2.42***	-2.44***	-2.35***
	Median	2.27***	2.49***	2.16***
	75th pct.	17.24***	19.00***	15.28***
	Std. dev.	13.31***	13.95***	13.01***++
	Skewness	0.42***	0.32***	0.46***
Hours worked per week				
	Mean	-0.04***	-0.02*	-0.05***++
	25th pct.	-0.14***	-0.14***	-0.15***
	Median	-0.01	-0.01	-0.00
	75th pct.	0.09***	0.10***	0.08***
	Std. dev.	0.30***	0.32***	0.29***++
	Skewness	0.13	0.49***	-0.11+++
Night shifts per month				
	Mean	-0.74***	-0.83***	-0.70***
	25th pct.	-1.17***	-1.36***	-1.10***+
	Median	-0.31***	-0.32***	-0.30***
	75th pct.	0.16***	0.15***	0.16***
	Std. dev.	1.86***	1.94***	1.81***
	Skewness	-1.01***	-1.11***	-0.94***
Predictable schedule				
	Mean	1.99***	2.46***	1.78***+++
	25th pct.	0.12*	0.22***	0.04++
	Median	1.13***	1.37***	1.04***+++
	75th pct.	3.40***	4.14***	3.00***+++
	Std. dev.	3.34***	3.82***	3.07***+++
	Skewness	0.98***	0.90***	0.93***
% rise in earnings after 10 years				
	Mean	0.09***	0.09***	0.09***
	25th pct.	0.01***	0.01***	0.01***
	Median	0.05***	0.05***	0.05***
	75th pct.	0.16***	0.15***	0.16***
	Std. dev.	0.11***	0.11***	0.11***
	Skewness	0.84***	0.86***	0.82***
Commuting time (min)				
	Mean	0.01	0.03***	0.00++
	25th pct.	-0.16***	-0.16***	-0.15***
	Median	0.02**	0.03***	0.02*
	75th pct.	0.21***	0.25***	0.19***
	Std. dev.	0.32***	0.32***	0.32***
	Skewness	-0.12**	-0.01	-0.17***++
Share of female colleagues				
	Mean	0.00	0.01***	0.00+++
	25th pct.	-0.04***	-0.03***	-0.04***+++
	Median	0.00*	0.01***	0.00++
	75th pct.	0.04***	0.05***	0.03***+++
	Std. dev.	0.09***	0.09***	0.09***
	Skewness	0.26*	0.35**	0.22
Nb. individuals		1,146	365	781

Notes: ***, **, * denote significance based on bootstrap standard errors at the 1%, 5% and 10% levels, respectively. We also test for differences between genders. +++, ++, + denote significance of these tests at the 1%, 5% and 10% levels, respectively.

Table E2: Descriptive Statistics on the WTPs Estimated from the Job Choice Model.

		Overall	Men	Women
Hours worked per week				
	Mean	2.36***	1.94***	2.55***+
	25th pct.	0.24**	0.24**	0.24**
	Median	1.78***	1.55***	1.84***++
	75th pct.	3.33***	2.92***	3.61***++
	Std. dev.	6.31***	6.92***	5.99***
	Skewness	2.14**	1.03	2.95***
Night shifts per month				
	Mean	6.15**	11.50***	3.66+
	25th pct.	-11.77***	-11.24***	-11.77***
	Median	0.57	0.65	0.53
	75th pct.	10.23***	9.16***	11.14***
	Std. dev.	87.69**	96.77***	83.05**
	Skewness	14.64**	7.01	20.09***
Predictable schedule				
	Mean	175.70***	229.39***	150.61***
	25th pct.	-18.62***	-17.48***	-18.94***
	Median	-1.17	-0.80	-1.38
	75th pct.	39.55***	44.03***	34.44***
	Std. dev.	1,735.85***	2,646.39***	1,073.77***+++
	Skewness	21.42***	17.36***	11.32***++
% rise in earnings after 10 years				
	Mean	-0.20**	-0.31***	-0.15*
	25th pct.	-1.05***	-1.13***	-1.02***++
	Median	-0.52***	-0.54***	-0.52***
	75th pct.	0.39***	0.22**	0.49***++
	Std. dev.	2.48***	2.56***	2.45***
	Skewness	1.09	1.07	1.10
Commuting time (min)				
	Mean	3.37***	3.08***	3.51***
	25th pct.	0.71***	0.56***	0.81***
	Median	2.13***	1.90***	2.30***++
	75th pct.	4.78***	4.75***	4.78***
	Std. dev.	7.69***	7.33***	7.85***
	Skewness	2.11***	1.98***	2.15***
Share of female colleagues				
	Mean	-0.13	0.05	-0.21**++
	25th pct.	-0.58***	-0.50***	-0.66***
	Median	0.02	0.03	0.02
	75th pct.	0.49***	0.50***	0.48***
	Std. dev.	2.57***	2.46***	2.61***
	Skewness	-0.81	-1.22**	-0.64
Nb. individuals		1,146	365	781

Notes: ***, **, * denote significance based on bootstrap standard errors at the 1%, 5% and 10% levels, respectively. We also test for differences between genders. +++, ++, + denote significance of these tests at the 1%, 5% and 10% levels, respectively.

E.3.2 Source of Heterogeneity in Estimated Preferences

Wiswall and Zafar (2018) highlights two potential sources of heterogeneity in estimated preferences: true heterogeneity in preferences for job attributes, and differences in the respondents' perceptions of the hypothetical jobs' unspecified dimensions. In this section, we examine the extent to which observable characteristics can explain the underlying heterogeneity in WTP estimates. We do this by regressing the estimated WTP on a list of individual covariates using multivariate linear regressions. In particular, we include gender, age, an indicator for having the French nationality, fertility plans for the next five years, relationship status, parental education, and NRE rank quartiles. The results are presented in Table E3.

Table E3: WTP Estimates and Individual Characteristics.

	(1) Hours	(2) Night shifts	(3) Pred. schedule	(4) % rise	(5) Commute	(6) Prop. women
Woman	0.480	-6.211	53.50	0.196	0.337	-0.393**
	(0.439)	(4.589)	(61.28)	(0.173)	(0.503)	(0.158)
Age	0.0156	0.974	-4.124	-0.0399	-0.108	-0.0899*
	(0.0928)	(1.012)	(11.38)	(0.0563)	(0.153)	(0.0499)
French nationality	0.917	-1.427	87.54	0.627**	0.871	-0.388
	(1.263)	(9.542)	(64.60)	(0.266)	(1.067)	(0.526)
Fertility plans						
Does not know	1.084**	-12.50**	117.3	0.0282	0.778	0.0661
	(0.468)	(3.724)	(83.80)	(0.188)	(0.595)	(0.191)
Yes	0.902*	-7.814	7.703	0.0666	-0.0452	0.391**
	(0.523)	(4.917)	(50.96)	(0.189)	(0.551)	(0.176)
Relationship status						
In a relationship	0.146	-6.540*	18.14	0.183	0.0659	0.157
	(0.442)	(3.345)	(62.80)	(0.160)	(0.491)	(0.161)
Married	-0.0780	-7.258	-10.81	0.443	1.961	0.952**
	(1.010)	(6.201)	(80.68)	(0.583)	(1.201)	(0.463)
Parents' education						
One higher education	0.306	-3.230	-139.6	0.234	1.013	0.114
	(0.600)	(6.101)	(119.8)	(0.232)	(0.704)	(0.223)
Both higher education	0.191	-5.087	-152.6	0.358*	0.543	-0.148
	(0.473)	(5.410)	(104.4)	(0.199)	(0.572)	(0.201)
Exam score quartiles						
Q_2	-0.118	-0.317	77.66	0.319	0.000738	-0.0114
	(0.682)	(7.037)	(65.13)	(0.268)	(0.725)	(0.261)
Q_3	-0.141	-7.628	150.4*	0.162	1.424*	0.385
	(0.690)	(6.091)	(77.81)	(0.260)	(0.788)	(0.260)
Q_4	0.158	-5.652	140.8**	0.224	0.653	0.219
	(0.751)	(6.448)	(60.60)	(0.253)	(0.712)	(0.255)
R^2	0.009	0.026	0.010	0.012	0.014	0.023
Observations	991	991	991	991	991	991

Notes: This table presents correlations between estimated WTP for different job attributes and respondents' individual characteristics. Sample used correspond to those respondents for whom we have all the included characteristics. Robust standard errors are shown in parentheses. *, ** and *** denote significance at the 1%, 5%, and 10% levels, respectively.

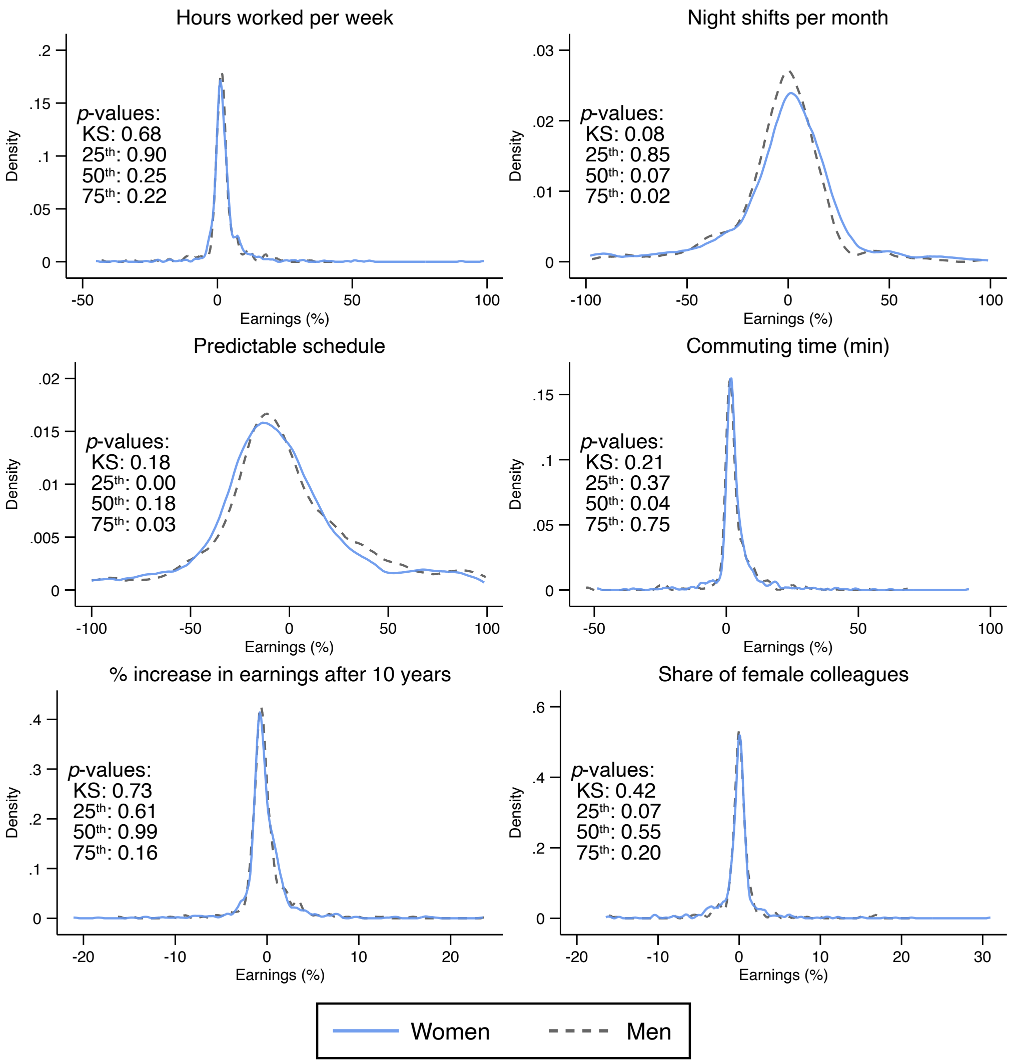
Two observations emerge from Table E3. First, none of the regression R-squared exceed 3 percent, which indicates that the included covariates only explain a very small fraction of the variation in WTP estimates. Second, most of the correlations are not statistically significant, which suggests that the observed variation in WTP estimates is not "controlled for" by these covariates.[63] Similar to Wiswall and Zafar (2018), these two observations taken together suggest that the observed heterogeneity in WTP estimates cannot be explained by basic demographic characteristics. They are likely to come from true underlying heterogeneity in preferences, rather than from differences in how respondents with different characteristics perceive and respond to the hypothetical job choice experiment.

E.3.3 Additional Heterogeneity Analyses

Figures E4 to E6 replicate Figure 4 using alternative specifications of equation (4). They show that our results are robust to changes in the specification of the model, specifically to the introduction of quadratic and interaction terms.

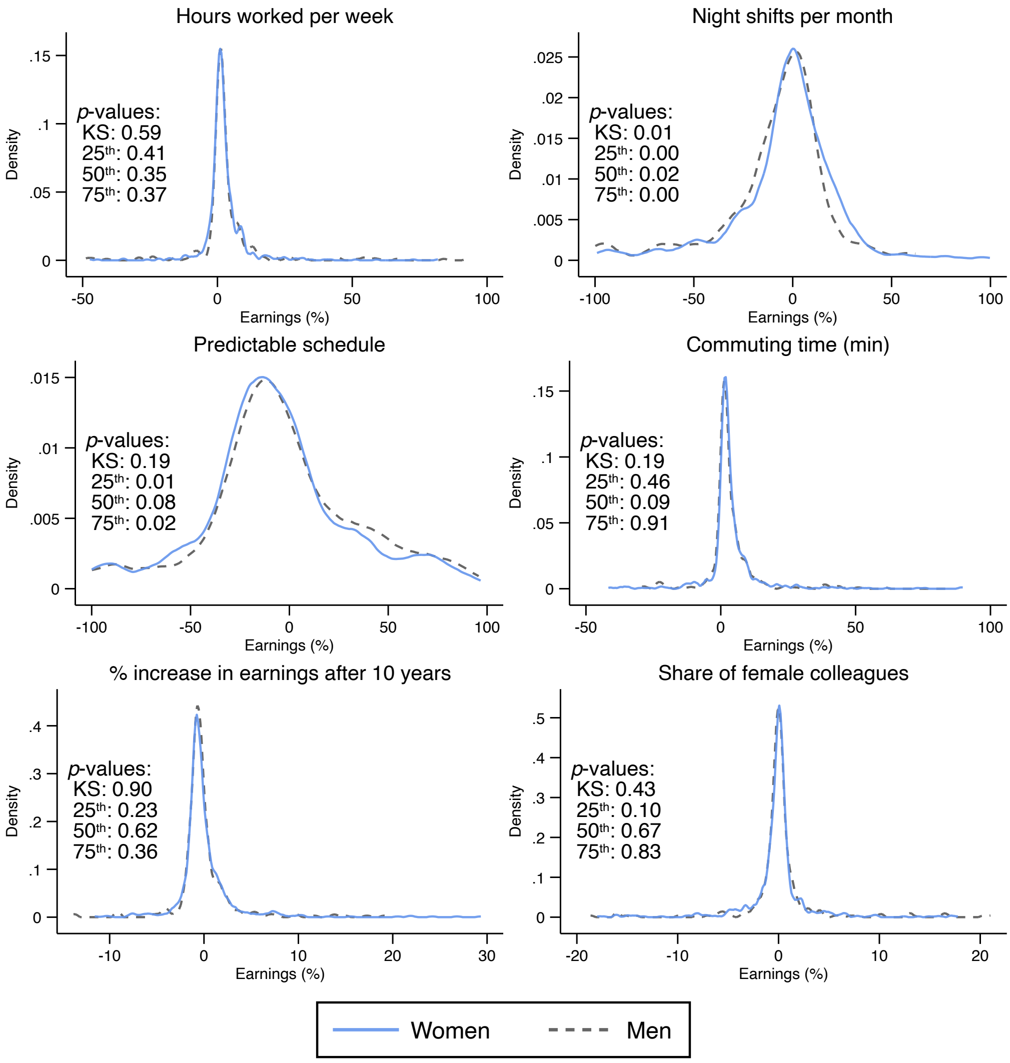
[63]Note that the lack of significance of the estimates on the gender dummy variable only indicates that there are no statistically significant gender differences in *mean* WTP for a given attribute. We have previously shown that it hides important heterogeneity.

Figure E4: Kernel Densities of WTP Estimates from Model with Quadratic Term for Hours Worked.



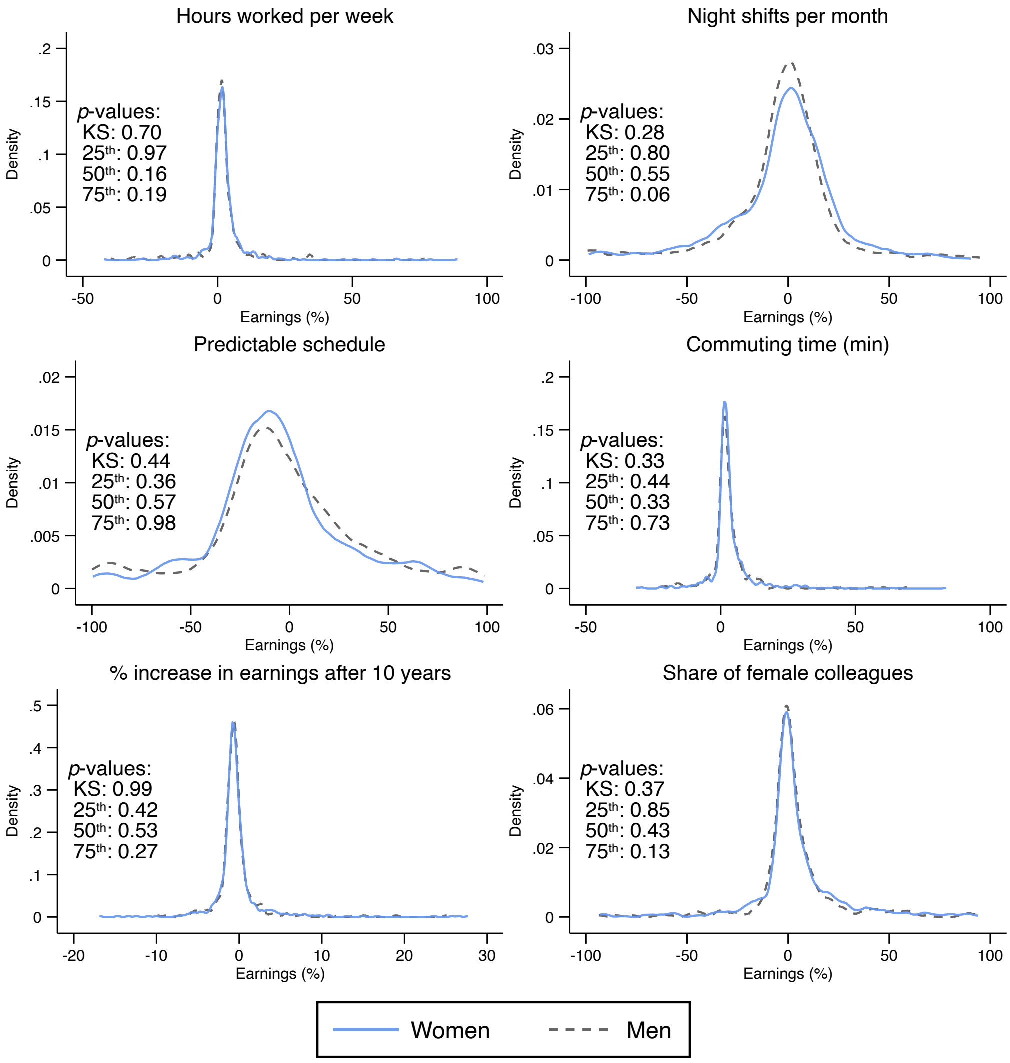
Notes: This figure displays the kernel densities of the percentage an individual needs to be compensated for a unit change in the job attribute or willingness to pay (WTP) computed using the relevant utility weight estimated from a version of equation (4) including a quadratic term for hours worked by gender. p-values for Kolmogorov-Smirnov tests of the equality of the two distributions. We also conduct tests for gender differences in the 25^{th}, 50^{th} and 75^{th} percentiles using bootstrapped samples. WTP values above (below) 100 (-100) are not shown.

Figure E5: Kernel Densities of WTP Estimates from Model with Interaction Term between Hours and Schedule Predictability.



Notes: This figure displays the kernel densities of the percentage an individual needs to be compensated for a unit change in the job attribute or willingness to pay (WTP) computed using the relevant utility weight estimated from a version of equation (4) including an interaction term between hours worked and schedule predictability by gender. *p*-values for Kolmogorov-Smirnov tests of the equality of the two distributions. We also conduct tests for gender differences in the 25th, 50th and 75th percentiles using bootstrapped samples. WTP values above (below) 100 (-100) are not shown.

Figure E6: Kernel Densities of WTP Estimates from Model with Quadratic Term for Share of Female Colleagues.



Notes: This figure displays the kernel densities of the percentage an individual needs to be compensated for a unit change in the job attribute or willingness to pay (WTP) computed using the relevant utility weight estimated from a version of equation (4) including a quadratic term for share of female colleagues by gender. p-values for Kolmogorov-Smirnov tests of the equality of the two distributions. We also conduct tests for gender differences in the 25^{th}, 50^{th} and 75^{th} percentiles using bootstrapped samples. WTP values above (below) 100 (-100) are not shown.

E.3.4 Willingness to Pay and Workplace Characteristics

To analyse whether preferences translate into actual choices, this section examines whether the estimated individual preferences for job attributes are related to the characteristics of the medical specialties that respondents actually choose for their residency training. We do this by relating each respondent's estimated willingness to pay for a job attribute to the job attribute's mean value in the specialty of choice. As noted above, the centralised allocation mechanism of students to medical specialties is incentive-compatible and minimises labour market frictions. By eliciting individual preferences for job attributes just weeks before one's actual specialty choice, we exploit an ideal setting in which to relate stated and revealed preferences.

Although we know the specialty chosen by each respondent for their residency training, we do not have information on actual job characteristics for each respondent. Therefore, one caveat of our data is that is the lack of individual-level measures of job attributes. We address this limitation by using the specialty-level characteristics presented in Section 2.2 on the number of hours worked, the number of night shifts performed, and the proportion of women in the specialty. After collapsing individual WTP estimates at the specialty level, we are left with 44 observations.

Table E4: Job Attributes of Actual Chosen Specialties and WTP.

	(1)	(2)	(3)
	Hours worked	Night shifts	Prop. women
WTP	-0.924	-0.008	-0.054**
	(0.723)	(0.008)	(0.021)
Constant	57.838**	3.973**	0.483**
	(2.031)	(0.161)	(0.013)
Mean dep.var	58.84	3.85	0.46
Std. dep. var	10.26	1.19	0.18
N	44	44	44

Notes: This table analyses the relationship between the estimated preferences for job attributes and those of the medical specialties chosen by respondents for their residency. The shown coeficients come from regressing the medical specialty mean value of the job attribute chosen by respondents for their residency on the mean value of the estimated willingness to pay (WTP) for that attribute. Estimated coeficients are weighted by the number of respondents who selected each specialty. Bootstrapped standard errors are shown in parentheses. *, ** and *** denote significance at the 1%, 5%, and 10% levels, respectively.

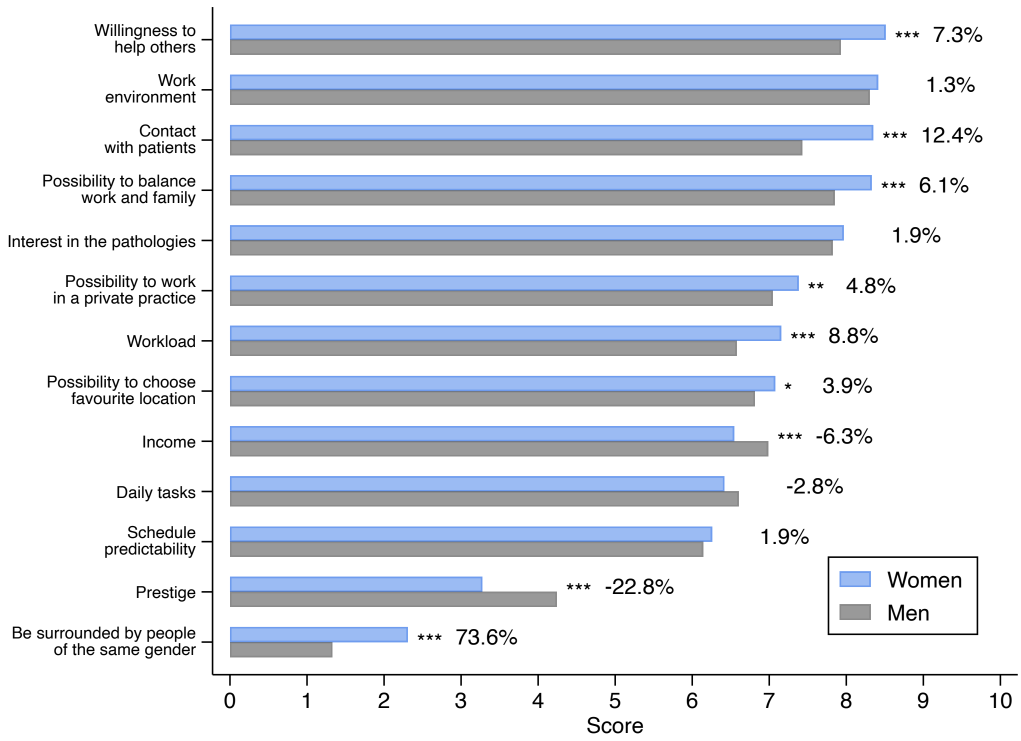
Recall that WTP as defined in equation (6) is the amount by which an individual needs to be compensated for a unit change in a given attribute. We thus expect each WTP to be negatively correlated with its corresponding job attribute. Table E4 reports the results obtained by regressing the mean value of a given job attribute on the mean value of the corresponding estimated WTP, and weighting the coefficients by the number of respondents who chose each specialty for their residency. All three coefficients are negative, and the relationship between WTP for a higher proportion of female colleagues is statistically significant at the 5 percent level. Note that this analysis includes the cases in which an individual does not get their preferred specialty and in which one might expect the relationship between estimated preferences and actual job attributes to be weakened. Even when these individuals are included, the negative expected relationships go through.

We also test for the existence of a relationship between estimated preferences and job characteristics within the individual. We follow Wiswall and Zafar (2018) and, for each attribute, rank the 44 specialties according to their mean estimated WTP and their attribute value. We then compute the individual-level correlation between ranked actual attributes and ranked WTPs, and weight each specialty by the number of respondents who actually chose it. We find a weighted mean correlation coefficient across all specialties of -0.30 which is statistically different from 0 at the 1 percent level, and a median correlation coefficient of -0.71, which is not statistically different from 0 at standard values.

These results suggest that our estimated preferences for job attributes do reflect the true underlying heterogeneity of preferences and relate to actual occupational choices. However, given their dispersion, some of our estimates appear not to measure preferences as accurately as we would have expected. In the next section, we complement the analysis performed in this section by looking at other measures of preferences for workplace characteristics.

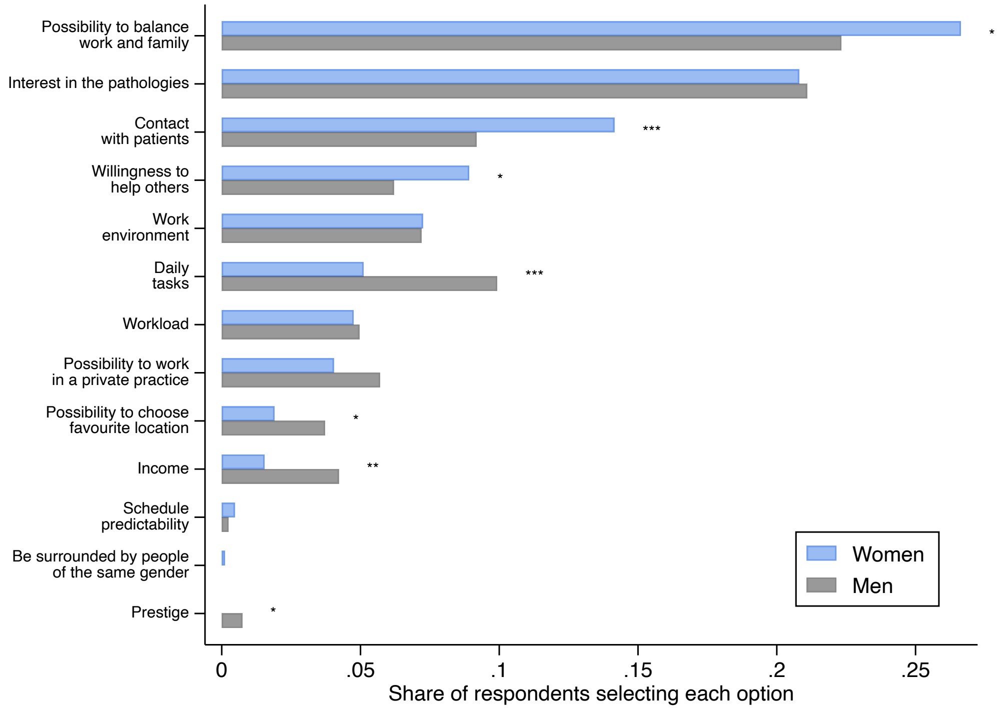
Appendix F Supplementary Material on Self-Reported Preferences

Figure F1: Self-Reported Valuation of Job Attributes by Gender.



Notes: This figure shows average self-reported tastes by gender for the 13 job characteristics that we asked respondents about in the survey via the following question: *"At the beginning of this questionnaire, you said that you would like to specialise in* specialty_preferred. *On a scale from 0 to 10, how important was each of the following factors in your specialty choice?"*. Significance levels of the two-sample t-tests for the difference in means between men and women are reported using *** p<0.01, ** p<0.05, * p<0.1. The percent difference between the female and male means is also provided for each characteristic.

Figure F2: Most Important Factor for Specialty Choice by Gender.



Notes: This figure shows the share of respondents who selected each of the presented features as the most important driver of their specialty choice. Significance levels of the two-sample t-tests for the difference in means between men and women are reported using *** p<0.01, ** p<0.05, * p<0.1.

F.1 Open-Ended Responses on the Attributes of Preferred Specialty

After indicating their preferred residency position, respondents were asked to provide positive (*pros*) and negative (*cons*) aspects of their preferred specialty using two free-text responses. This type of question allows respondents to express themselves freely about the attributes they believe are associated with their *preferred* specialty and are most important to them. It also allows us to identify the sentiment they have for each attribute, depending on whether it was mentioned as a *pro* or as a *con*.

Based on free-text responses, we define 32 attributes related to five broad categories, four of which are borrowed from Sockin (2021) and adapted to fit the French healthcare labour market.[64] The first category relates to pay and includes two attributes about base pay and its progression. The second category relates to working conditions and includes 17 attributes: work-life balance, work schedule, work intensity, commuting, teleworking, location, autonomy/responsibility, respect/abuse, difficulty, stress, safety, recognition, fun, diversity/inclusion, type of workplace, lack of emergency work, and emotional involvement.[65] The third category relates to human capital and includes career concerns, on-the-job training, and the medical residency. The fourth category is entitled interpersonal relationship and includes what relates to hierarchy, teams, patients, and whether the specialty has a social component. We add to Sockin's work and define a fifth category that is very relevant in the process of choosing a medical specialty. It relates to the work content of the job, and contains 6 attributes that capture whether the specialty involves technical, administrative, intellectual and preventive tasks, whether it is a multidisciplinary specialty, and finally whether it relates to women's healthcare.

In order to identify these attributes in the respondents' answers, we assign a set of words to each of them and search for these words in the free-text responses using a language recognition model.[66] Table F1 presents the words that were assigned to each amenity. We then computed the incidence rate of each amenity as a positive or a negative aspect of the respondents' preferred specialty, separately for men and women. We show these incidence rates in Figure F3.

The three attributes that are most frequently mentioned as *pros* by both men and women relate to the relationship with the patients (42 percent for women and 26 percent for men), the multidisciplinary nature of the job (32 and 36 percent) and the type of workplaces in which the job can be performed (31 and 23 percent). In addition, women are significantly more likely than men to mention contact with patients, the type of workplace, or the work schedule as a positive aspects of their preferred specialty, and they are significantly less likely than men to mention the technical nature of the job or the pay as

[64] The author analyses positive and negative attributes of job reviews posted by workers on a labour market website. The author implements a topic modelling machine learning algorithm to identify latent amenities in the text using topic-specific anchor words that help the algorithm's effectiveness in identifying topics that can be interpreted as specific amenities.

[65] For the details about the reasons behind the selection of all the other amenities we refer to the references mentioned in Sockin (2021).

[66] The free-text responses were pre-processed by removing common stop words and lemmatising each of the remaining words using the trained natural language recognition model *fr_dep_news_trf* developed with the software library *spaCy*.

Table F1: Classification and Identification of Amenities.

#	Category	Amenity	Searched words
1	Pay	Pay	paye, gagne, paie, rémunération, salaire, argent, bonus
2	Pay	Pay growth	augmentation annuelle, augmentation salariale, augmentation salaire
3	Working conditions	Work-life balance	vie, organisation personnelle, investissement
4	Working conditions	Work schedule	heure, horaire, temps, garde, astreinte, nuit, tôt, tard, chronophage, semaine
5	Working conditions	Work intensity	quantité travail, rythme, épuisement, prenant, vitesse
6	Working conditions	Commuting	trajet, stationnement, autobus, voiture
7	Working conditions	Teleworking	télétravail, téléconsultation, télé-consultation, travail domicile
8	Working conditions	Location	ville, lieu, localisation, urbain, rural
9	Working conditions	Workplace type	cabinet, libéral, hôpital, structure, salariat
10	Working conditions	Autonomy/responsability	autonomie, indépendance, responsabilité
11	Working conditions	Respect/abuse	dignité, abus, harcèlement, hostilité
12	Working conditions	Difficulty	défi, challenge, difficile, fastidieux, exiger
13	Working conditions	Stress	stress, pression
14	Working conditions	Safety	dangereux, sécurité
15	Working conditions	Recognition	travail acharné, effort, récompense, prestig, réputation, reconn, élitiste, valor, image, dénigr, opinion
16	Working conditions	Fun	amusant, ennuyeux, ennui, monotonie, monotone
17	Working conditions	Diversity/Inclusion	ethnique, multiculturel, inclusif, lgbtq, inclusion, égalité
18	Working conditions	No emergencies	pas d'urgence, peu d'urgence, sans urgence, pas trop d'urgence
19	Working conditions	Emotional involvment	psychologique, mental, décès, mort
20	Work content	Task technicity	technique, technicité, manuel, minutieux, minutie, chirurgicale, chirurgie, chir
21	Work content	Administrative tasks	administra, paperasse
22	Work content	Intellectual tasks	rélféxion, recherche, intellectuel, analyse
23	Work content	Preventive tasks	prévent, dépist, somatique
24	Work content	Multidisciplinary	pluridisciplin, polyvale, diversité pathologie, exercice mixte, transversal, multidisciplin
25	Work content	Female health	femme, santé femme, soigner femme, suivi femme, soin femme, acompagnement femme, contact femme
26	Human capital	Career concerns	carrière, croissance, amélioration
27	Human capital	On-the-job training	formation, surspécialisation, sur-spécialisation, sur spé, diplôme
28	Human capital	Medical residence	stagiaire, stage, internat
29	Relationships	Hierarchy	gestion activité, patron, chef service, hiérarchie
30	Relationships	Teams	équipe, collaboration, collègue, isole, seul, solitude
31	Relationships	Patient	patient, suivi long, contact, médecin famille
32	Relationships	Social component	humain, humanité, social, mission, valeurs, utilité, relation

Notes: This table lists the anchors that are used to identify each amenity in the survey respondents' free-text responses. These lists were constructed so as to have a balanced number of anchors across amenities. We excluded anchors that could have ambiguous applied to several amenities. Some of these anchors are root words, e.g. we use *prestig* to find both *prestigious* and *prestige* in the free-text responses.

positive aspects.[67] Interestingly, women are also more likely than men to care about working with female patients, and this holds both across and within specialty types.

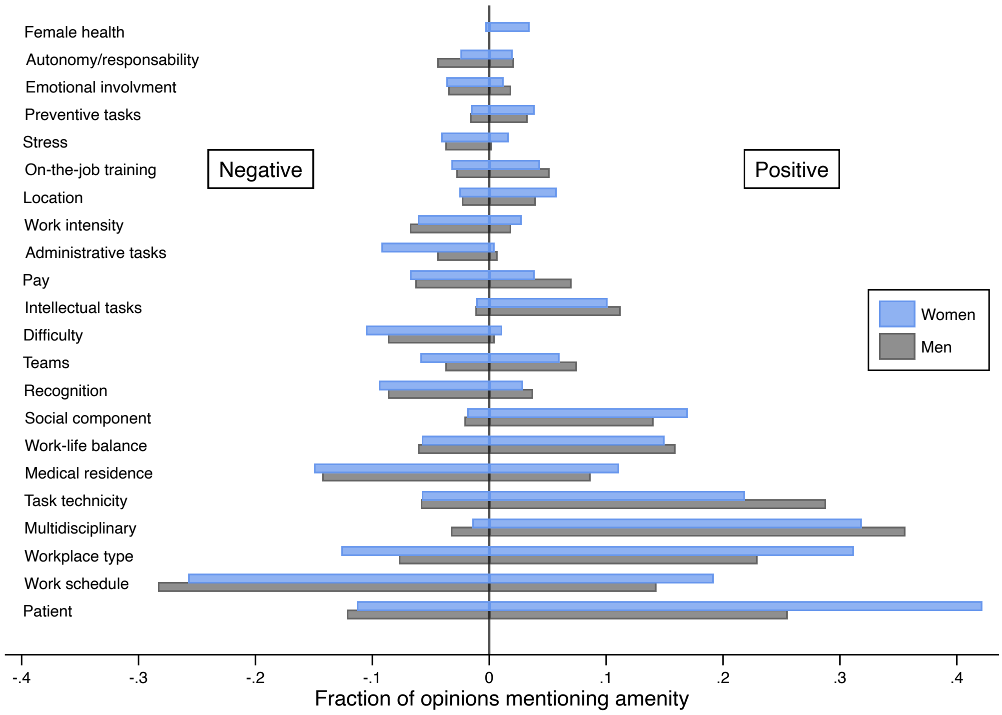
The job attribute that is by far the most frequently mentioned as a *con* by both men and women relates to the work schedule of the job (26 percent for women and 28 percent for men).[68]

Overall, these results confirm those found with different methodologies and question formats, namely that men and women, on average, differ in what they care about in a job. They also suggest that our previous analyses do a good job in identifying the job attributes that are relevant to medical students when choosing a specialty.

[67] Table F2 reports the significance levels of the tests of differences in means between men and women.

[68] The fact that some amenities are mentioned both as *pros* and *cons* reveals the existence of individual heterogeneity in opinions about the attributes that encourage and discourage choosing a specialty.

Figure F3: Positive and Negative Attributes Identified in Open-Ended Responses about One's Preferred Specialty.



Notes: This figure reports the fraction of times each attributes is identified in the free-text responses about the positive and negative attributes of the reported preferred speciality. Only amenities that are found in more than 3% of answers are shown. Words used to identify each attribute in the free-text are reported in Appendix D Table F1. Texts were pre-processed by removing common stop words in French and lemmatising each of the remaining words using the trained model for natural language recognition *fr_dep_news_trf* developed with the software library *spaCy*. Amenities listed in ascending order according to the incidence rate of both types of opinions provided by women.

Table F2: Prevalence of Each Amenity as a *Pro* or *Con*.

Category	Amenity	Positive characteristics (%)			Negative characteristics (%)		
		Women	Men	Difference	Women	Men	Difference
Pay	Pay	3.88	7.03	-3.14**	6.77	6.32	0.45
Pay	Pay growth	0.00	0.00	0.00	0.00	0.00	0.00
Working conditions	Work-life balance	14.98	15.93	-0.94	5.77	6.09	-0.32
Working conditions	Work schedule	19.20	14.29	4.92**	25.75	28.34	-2.59
Working conditions	Work intensity	2.77	1.87	0.90	6.10	6.79	-0.69
Working conditions	Commuting	0.00	0.00	0.00	0.00	0.00	0.00
Working conditions	Teleworking	0.00	0.00	0.00	0.00	0.00	0.00
Working conditions	Location	5.77	3.98	1.79	2.55	2.34	0.21
Working conditions	Workplace type	31.19	22.95	8.24***	12.65	7.73	4.92***
Working conditions	Autonomy/responsability	2.00	2.11	-0.11	2.44	4.45	-2.01*
Working conditions	Respect/abuse	0.11	0.23	-0.12	0.11	0.00	0.11
Working conditions	Difficulty	1.11	0.47	0.64	10.54	8.67	1.88
Working conditions	Stress	1.66	0.23	1.43***	4.11	3.75	0.36
Working conditions	Safety	0.11	0.23	-0.12	0.11	0.70	-0.59
Working conditions	Recognition	2.89	3.75	-0.86	9.43	8.67	0.77
Working conditions	Fun	0.55	0.00	0.55**	0.44	0.70	-0.26
Working conditions	Diversity/Inclusion	0.00	0.00	0.00	0.00	0.00	0.00
Working conditions	No emergencies	0.00	0.00	0.00	0.00	0.00	0.00
Working conditions	Emotional involvment	1.22	1.87	-0.65	3.66	3.51	0.15
Work content	Task technicity	21.86	28.81	-6.94***	5.77	5.85	-0.08
Work content	Administrative tasks	0.44	0.70	-0.26	9.21	4.45	4.76***
Work content	Intellectual tasks	10.10	11.24	-1.14	1.11	1.17	-0.06
Work content	Preventive tasks	3.88	3.28	0.61	1.55	1.64	-0.09
Work content	Multidisciplinary	31.85	35.60	-3.74	1.44	3.28	-1.84*
Work content	Female health	3.44	0.00	3.44***	0.33	0.00	0.33*
Human capital	Career concerns	2.22	2.11	0.11	1.22	0.94	0.28
Human capital	On-the-job training	4.33	5.15	-0.82	3.22	2.81	0.41
Human capital	Medical residence	11.10	8.67	2.43	14.98	14.29	0.70
Relationships	Hierarchy	0.78	0.23	0.54	0.00	0.00	0.00
Relationships	Teams	5.99	7.49	-1.50	5.88	3.75	2.14*
Relationships	Patient	42.18	25.53	16.65***	11.32	12.18	-0.86
Relationships	Social component	16.98	14.05	2.93	1.89	2.11	-0.22
Observations		899	427		896	424	

Notes: This table shows the prevalence in the respondents' free-text responses of each amenity defined in Table F1 as a positive or a negative characteristic, separately by gender. Texts were pre-processed by removing common stop words in French and lemmatising each of the remaining words using the trained model for natural language recognition *fr_dep_news_trf* developed with the software library *spaCy*.