The Un'Healthy' Gaps: Evidence on Gendered Faultlines in Digital Healthcare Services*

Dweepobotee Brahma† Nikita Sangwan‡

August 30, 2025

Preliminary draft - Please do not cite or circulate

Abstract

This paper examines gender disparities in engagement on a large-scale digital health platform in a developing country context. Using a novel high-frequency dataset from India covering over 7,000 physicians between September to December 2024, we explore differences in labor supply, patient engagement, pricing, and platform visibility by the physician's gender. We find that despite no difference in labour supply or consultation fees, female doctors face lower booking rates and are ranked lower in platform search results. To better understand the mechanisms behind these disparities, we conducted a laboratory experiment that reveals evidence of user's taste-based bias and helps tease out this effect from bias stemming from the platform's ranking algorithm. The presence of user bias is further substantiated using textual analysis of patient reviews using Natural Language Processing (NLP) tools from the Machine Learning literature. We find marginally lower recommendation rates observed for female physicians even when the overall volume of reviews is similar across genders when controlling for patient volume. Our findings suggest that digital health platforms may reproduce traditional gender-based inequalities even in high-skilled professional contexts, raising important questions about algorithmic visibility and implicit bias.

JEL classification: J16, J24, I11, O33, L86
Keywords: gender bias, digital platforms, healthcare, high-skilled professionals

*We thank the attendees at the Queen's University Belfast Seminar, Global Development Institute (Manchester) Seminar, Liverpool Management School Seminar and IHEA Conference for helpful comments. We acknowledge excellent research assistance from Khushal Sindhav, Abhijeet Balaji Chavan, Deepti, and Cara Taggart. The authors acknowledge the financial support from the Google India Research Award 2021. The usual disclaimers apply.

†Indian Institute of Technology, Jodhpur. Email: dweepobotee@iitj.ac.in

‡Queen's Business School. Queen's University Belfast. Email: nikitasagwan24@gmail.com

1 Introduction

Digital health platforms have significantly reshaped healthcare delivery, particularly in low- and middle-income countries where access to care remains limited. Traditional healthcare labor markets are characterized by persistent gender gaps in opportunities and earnings among practitioners. In contrast, digital markets introduce features such as flexible work arrangements, shorter engagements, and reputation systems. These features could help narrow gender disparities in healthcare provision, but they also carry the risk of reinforcing them. Despite the rapid growth of digital platforms, relatively little is known about their implications for healthcare labor markets specifically, or more broadly for gender inequalities in labor markets mediated by digital platforms. This paper examines the labor market for healthcare service providers operating through digital health platforms.

In this study, we investigate whether digital platforms offer a more equitable space for participation or whether gender disparities seen in traditional settings persist even in digital environments. Specifically, this study examines the gendered engagement of healthcare service providers on a large-scale digital health platform in a large developing country context, India. Using a novel high-frequency dataset from a popular digital health platform in India, we examine gender differences in labor supply, patient engagement, pricing, and platform visibility by doctor's gender. We find that despite no difference in labour supply by physicians or in consultation fees, female doctors experience systematically worse outcomes: they have lower booking rates, are ranked lower on the platform, and receive fewer patient ratings as well as lower recommendation rates compared to their male counterparts. The observational data from the platform alone cannot disentangle the mechanisms driving lower booking rates - whether these stem primarily from reduced visibility due to lower ranks, or from fewer patient ratings and less favorable recommendations, which themselves may influence rank. To address this, we conducted a laboratory experiment designed to isolate these channels on a similar simulated digital health platform. The experiment reveals that ranking has no significant effect on booking rates. Instead, users exhibit a stronger preference for male physicians, a bias that likely contributes directly to the observed gender gap in bookings. Using this knowledge, we return to the data from the digital platform and use the text data from the patient reviews to uncover the pathway in which users' bias manifests into lower recommendations for female doctors.

This paper speaks to several streams of the literature. First, while this paper relates to the larger literature on gender gaps in the labour market, it specifically contributes to the stream on gender gaps on digital platforms. Prior studies have documented the existence of gender gap on the digital platform through the gig economy. Cook *et al.* (2020) find a

7% gender earnings gap among drivers on the Uber rideshare app arising from learning on the platform, locational constraints, and driving speed. Foong *et al.* (2018) study Upwork - a popular gig work platform and find that the median female worker on the platform set hourly wages that were 74% of the median man's hourly wage, - a difference that cannot be fully explained by experiences, education, or job category.

Second, this study contributes to the stream of research investigating gender gaps in physician outcomes.Gender gaps in earnings are attributed either to lower labour supply (Kehrer (1976),Ohsfeldt & Culler (1986) Sasser (2005), , productivity Langwell (1982), marriage or children Sasser (2005) , discrimmination Ohsfeldt & Culler (1986), Chen (2024), differential interpretation of quality signals Sarsons (2017). Beyond earnings, gender gap also exist through slower career progression and Ash *et al.* (2004) and lower rates of referrals Sarsons (2017). Existing research on gender gaps among physicians is largely focused on high-income countries (Baker (1996), Bashaw & Heywood (2001)Gravelle *et al.* (2011), Theurl & Winner (2011), Jena *et al.* (2016) Lo Sasso *et al.* (2020), Catenaccio *et al.* (2022)), with very limited evidence from low- and middle-income settings such as China (Chen (2024)). This is the first paper to examine gender disparities on a digital health platform in India.

Third, we build a unique, granular panel dataset of over 50,000 observation of over 7,000 unique doctors from a popular digital health platform in India that allows us to observe physicians' labour supply at both the extensive (whether they participate) and intensive (how much they work) margins. This enables us to analyse short-term gender dynamics in work patterns, which are often overlooked in studies relying on annual income data. Finally, we complement our quantitative analysis with a textual analysis of patient reviews using natural language processing (NLP) tools from the Machine Learning literature. This allows us to document gender-based differences in patient sentiments and satisfaction. To further disentangle the drivers of observed disparities, we conduct a small-scale laboratory experiment, which reveals evidence of taste-based discrimination from users. We also use Natural Language Processing tools to analyze patient reviews and corroborate the sentiment patterns. To our knowledge, this is among the first applications of NLP in the literature on health economics to study gender bias using patient-generated text data. To our knowledge, this is also the first paper to combine secondary data analysis, a laboratory experiment and textual analysis to build a comprehensive body of evidence on the nature and mechanism and gender gaps.

In this paper, we find that despite having the same availability on the platform and charging similar prices, female doctors have significantly lower booking rates compared to male doctors. Female doctors also receive fewer patient reviews and lower recommendations compared to their male counterparts, and are ranked much lower on the search results of the platform for a given city within a specialization. However, the secondary data from the platform does not allow us to determine the mechanism behind the gender-gap in booking rates. We conduct a laboratory experiment on a simulated digital health platform with fictitious doctor profiles created using Generative AI and find evidence of taste-based discrimination against female doctors. We explore several mechanisms behind the source of this discrimination and find the gender homophily, education and prior beliefs drive it. We return to the platform data and analyze the text reviews using Natural Language Processing tools to pin-point how this taste-based discrimination manifests itself through lower recommendations to female doctors plausibly resulting in a lower booking rate.

The rest of the paper is organized as follows. Section 2 provides a background of the digital health platform and its functionality, and the market for physicians in India. Section 3 outlines our data scraping process and describes our sample. Section 4 presents our empirical strategy for the secondary data on the digital platform. Section 5 presents results across the secondary data from the platform, the laboratory experiment and the textual analysis of patient reviews. Section 6 concludes with a discussion.

2 Background

2.1 About the platform

We use data from a leading digital health platform with pan-India operations. The platform functions as an online marketplace for healthcare services, analogous to product search on an e-commerce website. Patients initiate a search by entering their city and required medical specialization, which generates a ranked, scrollable list of doctors. Each doctor is presented in a card format that provides key details, including the doctor's name, specialization (e.g., General Physician), subspecialty expertise (e.g., Diabetology within General Medicine), years of clinical experience, clinic location, hospital affiliations, and consultation fee. The card also reports aggregate ratings and the number of patient reviews. [1] Selecting a doctor's card allows patients to view the full schedule of available consultation slots for the upcoming days (both online and offline) and to access detailed patient feedback. Following a consultation, patients are invited to submit reviews and recommendations. These submissions are verified by the platform to remove spam, advertisements, and inappropriate content before being published on the physician's profile. Based on this feedback, the platform constructs a

[1] The platform also verifies the doctors before listing their profiles and displays a verified badge on the profile.

"recommendation score", defined as the percentage of all recommendations that are positive.[2]

2.2 Private Practice Market for Doctors in India

In India, physicians complete a five-year undergraduate degree in Medicine (MBBS) or Dental Surgery (BDS), followed by optional postgraduate specialization such as a Doctor of Medicine (MD) or Master of Dental Surgery (MDS), typically of three years duration. Upon graduation, physicians may practice in either the public or private sector. The public sector offers greater job security, with physicians generally receiving a fixed salary independent of patient volume. In contrast, the private sector operates largely on a fee-for-service model, where physician earnings depend on the number of consultations. Specialists, particularly surgeons, often affiliate with private hospitals while also maintaining independent clinics, thereby dividing their time across multiple practice settings. Owing to this structure, private-sector practice generally provides greater flexibility in both working hours and overall workload.

3 Data and Methodology

3.1 Data Collection Process

Our scraping tool was designed to extract detailed physician information, including a unique identifier, gender, specialization, years of experience, consultation fee, patient ratings (average rating and number of reviews), location (city and clinic locality), and consultation availability (days and hours). Data collection was carried out at the beginning of each week over four consecutive weeks, in two separate waves conducted in September 2024 and December 2024. For each physician, we recorded the total number of consultation slots available for the next seven days, providing a consistent weekly snapshot of both availability and booking patterns.

To calculate weekly availability, we first determined the duration of a consultation slot by measuring the time interval between two consecutive slots on the same day. The number of available slots per day was then multiplied by this duration, and the resulting values were aggregated across all seven days to obtain the total weekly availability in minutes. A similar procedure was followed for the booked slots, which yielded the total duration booked per week. Finally, we constructed the weekly booking rate for each physician, defined as the ratio of booked slots to total available slots.

[2]The recommendation score is displayed only once a physician has received at least ten verified recommendations.

3.2 Our sample

Table 1 provides the descriptive statistics of our sample. Column (1) displays the overall average of the variables in our sample, column (2) reports the average for male doctors, while column (3) reports the average for female doctors. Column (4) reports the raw difference in averages of the variables (men − women) and column (5) reports the standard errors of this difference. Our sample consists of 50,291 observations at the physician-week level. Of these 50.36% of the sample is constituted by female doctors while the remaining 49.64% are male physicians. Male physicians have close to 21.24 years of experience on average, while female doctors have a slightly lower average experience of 18.37 years. Male doctors offer an average of 72.33 slots per week for patient consultation through the platform, while female doctors offer a higher average of 76.94 slots per week. Similarly, the raw differences show that male doctors spend less time in patient consultations through the platform with an weekly average of 1546.19 minutes and compared to the female doctors' average of 1714.63 minutes per week. Female doctors also offer slightly longer slots than male doctors on average. Female physicians also charge a lower average consultation fee compared to male doctors yet have a lower booking rate. This result is in line with what Chen (2024) find in China. On the platform, female doctors have a lower number of reviews left behind by patients on average but a higher average recommendation rate. Female doctors are also ranked around 60 ranks lower on average compared to their male counterparts.

Our sample spans six specialties most commonly available on the platform - Cardiology, Dentistry, General Physician, Gynaecology, Orthopaedics and Pediatrics. The highest proportion of doctors on the platform are dentists followed by gynaecologists and general physicians. Dentistry has the highest proportion of female doctors relative to male doctors followed by gynaecology. In all other specialities, male doctors outnumber female doctors. Digital health platforms are largely prevalent in urban areas with very little penetration is smaller towns or rural areas. Our sample is spread over six major cities in India. Bangalore has the highest percentage of doctors listed on the platform at 35% followed by Delhi, Chennai, and Hyderabad at 16%, 16% and 15% respectively. Mumbai has a slightly lower share of doctors at 13% and Kolkata has the lowest share at 6%.

3.3 Lab Experiment

To examine the role of taste-based bias in patient choices on digital health platforms, we conducted a discrete choice experiment (DCE) at the end of November 2024. The experiment was administered to approximately 270 undergraduate students at the Indian Institute of

Technology (IIT) Jodhpur.[3] Participants interacted with a custom web interface designed to mimic the user interface of a real digital health platform. Each respondent was shown a series of fictitious physician profiles, with attributes systematically varied across choice sets. The attributes included physician gender (male or female), consultation fee (INR 600, 800, or 1000), years of experience (8, 15, 25, or 40 years), and patient rating (3, 4, or 5 stars). We created a full factorial design of all possible combinations and applied a D-efficient fractional factorial design to generate 15 choice sets. The summary statistics of the characteristics of the respondents of the survey are reported in Panel A of Table 2 . In each choice task, participants were presented with six doctor cards with a pair of doctors - one male and one female, displayed at the same rank with similar attributes to isolate gender effects. This design allows us to detect systematic preferences based on gender while holding other characteristics constant.

4 Empirical Strategy

Our preferred estimation equation is as follows:

$$y_{icts} = \alpha + \beta \, \text{Female}_{icts} + \gamma X_{icts} + \delta_c + \delta_t + \delta_s + u_{icts} \tag{1}$$

where y_{icts} is the outcome variable for doctor i in city c in wave t for specialization s. $Female_{icts}$ is the indicator for gender which takes the value of 1 if doctor i is female. The controls X_{icts} are the doctor's characteristics. δ_c are city fixed effects, δ_t are wave fixed effects, δ_s are specialization fixed effects, and u_{icts} is the error term. β is the parameter of interest capturing the estimated gender gap after controlling for physician location, qualifications, specialization, and wave. Thus, any remaining gender variation reflected in this parameter likely indicates bias - either stemming from user behavior on the platform or from the platform's own algorithmic design.

We structure our analysis around three core areas: outcomes on the platform service provider side (i.e., physicians), patient responses to these providers on the platform, and user feedback based on service delivery. Specifically, we examine six key variables of interest. First, we assess physician participation on the platform at both the extensive margin (number of consultation slots offered per week) and the intensive margin (total duration in minutes offered per week). We also assess the consultation fees set by doctors, shedding light on gender differences in pricing strategies. This captures the supply-side behaviour of doctors. Second, we analyse booking rates and the physician's rank on the platform to understand patient

[3] Ethics approval for the experiment was granted by the Institute Ethics committee at IIT Jodhpur

demand for these physicians. Finally, we investigate platform engagement and user feedback through the indicator of the number of reviews left by patients and the recommendation rate (i.e., share of patients who would recommend the physician). This final set of variables captures patient satisfaction with the healthcare services delivered by doctors and the visibility dynamics shaped by platform algorithms.

5 Results and Discussion

5.1 Platform data

Table 3 reports the estimates of gender gaps for our key variables of interest. Panel A reports the estimates of the gender gaps after incorporating practice locality, city, week, qualification, and specialization fixed effects. Once we account for the fixed effects, many of the differences between male and female doctors disappear. There is no statistical difference between male and female doctors in their availability on the platform either at the extensive margin through the number of available slots (column 1) or the intensive margin through the total duration of availability on the platform (column 2). Further, there is no statistical difference in the log consultation fee charged by male and female doctors (column 3). Despite this, there is a significant gender gap in booking rates between male and female doctors with female doctors having a booking rate of 0.014 lower than their male counterparts (column 4). Female doctors also receive on average 47 less reviews (column 5) compared to male physicians but without any significant gender gap in the recommendation percentage across these reviews (column 6). Additionally, female doctors are on average ranked 17 ranks lower than male doctors on the platform.

Panel B of Table 3 reports the estimates of the full regression after incorporating the doctor-specific controls for years of experience, a dummy for practice in a hospital, and rank on the platform. Even after the inclusion of doctor characteristics as controls, there is no significant gender difference in availability on the platform (columns 1 and 2) and price charged (column 3). However, female doctors continue to have a significantly lower booking rate compared to their male counterparts, although the gap shrinks to 0.011 upon the inclusion of controls. For consumer side outcomes, inclusion of controls reduces the gender gap in the number of reviews to 35.21 and reveals a statistically significant gender gap of 0.8% in the percentage of recommendation among reviews. The inclusion of doctor-specific controls maintains the significant gender gap in rankings and shrinks to 11.8.

The results of Table 3 beget the question of what drives a lower booking rate for female doctors relative to male doctors despite no differences in labour supply or consultation fees?

Is it purely a difference in the ranks on the platform? Or is it a difference in the consumer-side metrics such as reviews and recommendations, which in turn can drive the rank of doctors on the platform?

5.2 Lab Experiment

Secondary data from a digital health platform is insufficient in disentangling the two potential mechanisms driving lower booking rates for female doctors. Digital platforms could potentially be biased against female doctors and rank female doctors below male doctors. Patients could exhibit taste-based bias against female doctors and be less likely to choose female doctors. This could be reflected in the number of patient reviews and recommendation rates which in turn could make female doctors ranked lower than male doctors.

Panel B of Table 2 reports the willingness-to-pay (WTP) estimates for specific physician attributes on the platform. The baseline regression results from the DCE are presented in Appendix Table A.1. The most salient finding is that respondents are willing to pay INR 225 less to consult with a female physician, providing direct evidence of a monetary penalty associated with doctor's gender. By contrast, respondents are willing to pay INR 32 more for each additional year of experience and INR 716 more for each additional star in patient rating. These results suggest that while participants place positive value on physician experience and reputation, they also demonstrate a willingness to bear an economic cost to avoid female providers - evidence consistent with taste-based discrimination.

To further investigate the observed gender bias, we interact physician gender with several other variables, with results reported in Table 4. Columns 1 and 2 show that the interaction between physician gender and profile rank is statistically significant, indicating that presenting male and female doctors at the same rank does not eliminate the bias. This suggests that rank placement itself does not drive respondent choices. We also examine whether respondents' own characteristics mediate preferences by interacting physician and respondent attributes. Column 3 shows that female respondents are significantly more likely to choose female doctors, consistent with homophilic preferences or lower implicit bias. Moreover, respondents with higher educational qualifications (Masters/PhD) display weaker gendered preferences and rely more heavily on experience and ratings in their decision-making. Together, these findings highlight the importance of user demographics, particularly education, in shaping bias: more educated users appear better equipped to evaluate physicians on quality-related attributes rather than gendered stereotypes.

Finally, we leverage respondents' stated preferences to assess whether the observed bias stems from pre-existing beliefs and perceptions. In Table 4, Columns 5 - 7, we test for heterogeneity based on preferences for male physicians or perceptions of them as more experienced or skilled. Respondents who report a preference for male doctors are significantly less likely to select female physicians, and this bias persists even when female doctors possess comparable qualifications. While greater physician experience and higher ratings mitigate the bias to some extent, the overall effect remains substantial. Similar patterns emerge among respondents who perceive male doctors as more experienced or skilled. Taken together, these findings provide strong evidence of taste-based discrimination rooted in prior beliefs.

5.3 Textual Analysis

Having ruled out physician rank as the primary mechanism behind lower booking rates for female doctors, and having identified taste-based discrimination as a key channel, we return to the secondary data from the digital platform. Specifically, we analyze patient reviews, which form the basis of physicians' recommendation scores, to examine potential gender differences in review sentiment. To do so, we scraped a random sample of 165,826 text reviews from the platform. Of these, 82,734 reviews correspond to male doctors and 83,091 to female doctors. Because multiple reviews can be associated with each physician, this dataset spans 4,464 unique doctors, comprising 2,209 male and 2,255 female physicians.

We use VADER (Valence Aware Dictionary and Sentiment Reasoner) Hutto & Gilbert (2014) method to construct sentiment scores on the text reviews. VADER uses a predefined sentiment lexicon defined explicitly for sentiment analysis. This lexicon contains various words and phrases along with their corresponding sentiment ratings. To identify the sentiment in a given piece of text, VADER first breaks down the text into individual words. It then assigns a score to each word to identify whether it is positive or negative. Based on these scores VADER calculates the overall sentiment score of the text. The scores on every piece of text returned by VADER vary between -1 to +1, where -1 is very negative and +1 is very positive. We apply this sentiment analysis method to each review for each doctor and create the average sentiment score over all the reviews for each doctor.

We apply this sentiment analysis method to each review for each doctor and create the average sentiment score over all the reviews for each doctor. Appendix Table A.2 reports the average sentiment score over male and female physicians by specialization. The compound sentiment scores which combine both positive and negative sentiments for each text review is primarily positive across both genders and all specializations. This is also supported by the recommendation percentages data where the numbers are mostly in the ninenties. These compound sentiment scores are reported in column (1) while the positive sentiment scores are presented in column (2). The compound scores are very similar across male and female physicians for most specializations. The two exceptions are in the fields of orthopaedics and dentistry. Orthopaedics which has more male physicians than female physicians display a relatively higher lower compound score for female physicians than male physicians, while dentistry which has more female physicians than male physicians display a higher compound score for female dentists than male dentists. Similar differences are also observed in the positive compound scores across genders. Interestingly, gynaecology, which has a disproportionately higher number of female physicians compared to male physicians, has a relatively higher positive score for men than women. In Appendix Figure A.1, we plot the relative frequency of appearance of words identified as positive by VADER for male and female physicians. Most of the positive words (e.g. 'friendly', 'good', 'well', 'satisfied') appear less frequently for female physicians than for male physicians. We posit that the gender gap in recommendation percentage despite no significant difference in number of reviews stems from lower positive sentiments about the consultations by female doctors relative to male doctors.

6 Concluding Discussions

This paper examines gender disparities in outcomes on a digital health platform operating under a fee-for-service model. Using a novel, high-frequency dataset from a widely used platform in India, we investigate fine-grained differences in physician visibility and patient engagement. We find that, despite the absence of gender differences in labor supply (at either the extensive or intensive margin) or in consultation fees, female physicians experience systematically worse outcomes: they have lower booking rates, are ranked lower on the platform, and receive fewer patient ratings as well as lower recommendation rates compared to their male counterparts.

The observational data alone cannot disentangle the mechanisms driving lower booking rates - whether these stem primarily from reduced visibility due to lower ranks or from fewer and less favorable patient ratings and recommendations, which themselves may influence rank. To address this, we conducted a laboratory experiment designed to isolate these channels. The experiment shows that ranking has no significant effect on booking rates. Instead, users exhibit a stronger preference for male physicians, a bias that likely contributes directly to the observed gender gap in bookings. Complementing this, sentiment analysis of patient text reviews indicates that patients are marginally less likely to report positive experiences with female physicians. This subtle difference plausibly contributes to their lower recommendation rates. Together, these findings suggest that even in high-skilled digital labour markets, gender disparities persist and call attention to the need for more equitable design and governance of digital healthcare systems.

This study has several limitations. First, digital health platforms have limited reach beyond urban areas in India, and thus this data covers only six major cities in India. Second, the digital health platforms serve as directories for finding doctors in addition to allowing booking for consultations. Thus, we are not able to capture the part of the market that uses the platform to look up doctors but chooses to make bookings outside it. Unless these leakages are different across the genders, we do not believe they will affect our estimates.

While earlier papers find gender differences in labour supply among doctors, we do not. This provides more evidence towards the potential of digital platforms in lowering gender difference through increased flexibility. It is also important to note that in India, a large number of women study medicine often outnumbering men in certain specialties, e.g., dentistry.

References

Ash, A. S., Carr, P. L., Goldstein, R., & Friedman, R. H. 2004. Compensation and Advancement of Women in Academic Medicine: Is There Equity? *Annals of Internal Medicine*, **141**(3), 205–212.

Baker, L. C. 1996. Differences in Earnings between Male and Female Physicians. *New England Journal of Medicine*, **334**(15), 960–964.

Bashaw, David J., & Heywood, John S. 2001. The Gender Earnings Gap for US Physicians: Has Equality been Achieved? *LABOUR*, **15**(3), 371–391.

Catenaccio, Eva, Rochlin, Jonathan M., & Simon, Harold K. 2022. Addressing Gender-Based Disparities in Earning Potential in Academic Medicine. *JAMA Network Open*, **5**(2), e220067–e220067.

Chen, Yutong. 2024. Does the gig economy discriminate against women? Evidence from physicians in China. *Journal of Development Economics*, **169**, 103275.

Cook, Cody, Diamond, Rebecca, Hall, Jonathan V, List, John A, & Oyer, Paul. 2020. The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers. *The Review of Economic Studies*, **88**(5), 2210–2238.

Foong, Eureka, Vincent, Nicholas, Hecht, Brent, & Gerber, Elizabeth M. 2018. Women (Still) Ask For Less: Gender Differences in Hourly Rate in an Online Labor Marketplace. *Proc. ACM Hum.-Comput. Interact.*, **2**(CSCW).

Gravelle, Hugh, Hole, Arne Risa, & Santos, Rita. 2011. Measuring and testing for gender discrimination in physician pay: English family doctors. *Journal of Health Economics*, **30**(4), 660–674.

Hutto, C., & Gilbert, Eric. 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, **8**(1), 216–225.

Jena, Anupam B., Olenski, Andrew R., & Blumenthal, Daniel M. 2016. Sex Differences in Physician Salary in US Public Medical Schools. *JAMA Internal Medicine*, **176**(9), 1294–1304.

Kehrer, Barbara H. 1976. Factors Affecting the Incomes of Men and Women Physicians: An Exploratory Analysis. *The Journal of Human Resources*, **11**(4), 526–545.

Langwell, Kathryn M. 1982. Factors Affecting the Incomes of Men and Women Physicians: Further Explorations. *The Journal of Human Resources*, **17**(2), 261–275.

Lo Sasso, Anthony T., Armstrong, David, Forte, Gaetano, & Gerber, Susan E. 2020. Differences In Starting Pay For Male And Female Physicians Persist; Explanations For The Gender Gap Remain Elusive. *Health Affairs*, **39**(2), 256–263. PMID: 31967925.

Ohsfeldt, Robert L., & Culler, Steven D. 1986. Differences in income between male and female physicians. *Journal of Health Economics*, **5**(4), 335–346.

Sarsons, Heather. 2017. Interpreting signals in the labor market: evidence from medical referrals. *Job Market Paper*.

Sasser, Alicia C. 2005. Gender Differences in Physician Pay. *Journal of Human Resources*, **40**(2), 477–504.

Theurl, Engelbert, & Winner, Hannes. 2011. The maleâfemale gap in physician earnings: evidence from a public health insurance system. *Health Economics*, **20**(10), 1184–1200.

Table 1: Descriptive Statistics of the Characteristics of the Physicians

	All (N=50291)	Male (N=25048)	Female (N=25243)	Difference (Male-Female)
	(1)	(2)	(3)	(4)
Panel A: Descriptive Statistics of the Characteristics of the Physicians				
Experience (years)	19.80	21.24	18.37	2.87***
Hospital (ref: Clinic)	0.42	0.46	0.38	0.08***
Available Slots	74.65	72.33	76.94	-4.61***
Available Duration (mins)	1630.74	1546.19	1714.63	-168.44***
Booking Rate	0.06	0.06	0.05	0.01***
Consultation Fees	737.89	779.62	696.49	83.13***
Recommendation (%)	94.24	93.70	94.78	-1.08***
Rank	162.03	131.52	192.30	-60.78***
Reviews	71.30	78.62	64.15	14.47***
Qualifications				
BDS	0.32	0.28	0.36	-0.08***
MBBS	0.56	0.59	0.53	0.06***
MD	0.06	0.07	0.05	0.02***
MDS	0.04	0.04	0.03	0.01***
Panel B: Specialization Composition of the Physicians in our Sample				
Cardiologist	0.05	0.09	0.01	
Dentist	0.37	0.33	0.41	
General Physician	0.15	0.20	0.10	
Gynaecologist	0.21	0.03	0.39	
Orthopaedist	0.11	0.22	0.00	
Paediatrician	0.11	0.12	0.10	
Panel C: City Composition of the Physicians in our Sample				
Bangalore	0.35	0.32	0.38	
Chennai	0.16	0.18	0.14	
Delhi	0.16	0.16	0.16	
Hyderabad	0.15	0.14	0.15	
Kolkata	0.06	0.08	0.04	
Mumbai	0.13	0.12	0.14	

Note: Consultation Fees are measured in INR. Rank is a continuous variable capturing the overall rank of the physician within that city and specialization. Recommendation is the percentage of all recommendations that are positive. "Hospital" is a dummy equal to 1 if the physician is working at a hospitalwith the reference category being a clinic. Available slots are the total number of slots over the week. Available duration is the total minutes over the week. Booking rate is calculated as the total booked slots over total available slots for the upcoming week at the time of scraping the data. The qualification dummies equal 1 if that degree is the terminal degree of the physician. Column (4) in panel A reports the difference between male and female physicians and column (5) provides the corresponding standard errors. We test the difference between male and female physicians with a t-test of equal variance. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2: Summary Statistics of Experiment Respondents

Panel A: Summary Statistics of Experiment Respondents			
Variable	**Count**	**Mean**	**SD**
Age	265	20.079	2.815
Female respondent	269	0.130	0.337
Male respondent	269	0.862	0.345
Bachelor's degree (respondent)	269	0.807	0.396
Master's/PhD (respondent)	269	0.193	0.396
Familiar with doctor	267	2.494	1.249
Family doctor	269	0.335	0.473
Father as doctor	267	0.030	0.171
Mother as doctor	267	0.022	0.148
Brother as doctor	267	0.075	0.264
Sister as doctor	267	0.060	0.238
Other male member as doctor	267	0.202	0.402
Other female member as doctor	267	0.082	0.275
Last consulted doctor female	269	0.275	0.447
Last consulted doctor male	269	0.710	0.455
Male doctor perceived more experienced	267	0.494	0.501
Male doctor perceived more skilled	267	0.408	0.492
Preference for male doctor	267	0.404	0.492
Panel B: Estimates of Willingness to Pay from the Choice Experiment			
	Female	Experience	Rating
WTP	-225.66	31.93	715.78
Lower Limit (LL)	-412.10	14.20	339.13
Upper Limit (UL)	-39.22	49.66	1092.44

Table 3: Gender Gaps and Regression Results Across Outcomes

	Available Slots	**Available Duration**	**Log Consultation Fees**	**Booking Rate**	**Reviews**	**Recommendation**	**Rank**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Panel A: Regression Results with only Fixed Effects							
Female	-2.681	25.439	-0.010	-0.014***	-47.006***	-0.144	17.581***
	(1.752)	(33.287)	(0.024)	(0.003)	(5.394)	(0.333)	(3.529)
Observations	50,286	50,286	50,286	50,286	50,286	40,247	50,286
R-squared	0.298	0.374	0.397	0.127	0.136	0.185	0.498
Panel B: Regression Results with Additional Controls							
Female	-2.158	4.038	0.031	-0.011***	-35.216***	-0.841**	11.891***
	(1.676)	(32.185)	(0.025)	(0.004)	(4.605)	(0.331)	(3.513)
Hospital	-8.442**	-145.836**	0.180***	0.017***	-39.834***	-2.254***	8.923
	(3.342)	(68.184)	(0.045)	(0.006)	(6.967)	(0.563)	(7.493)
Experience	-0.470***	-16.873***	0.010***	0.000	2.220***	-0.225***	-1.842***
	(0.099)	(1.875)	(0.001)	(0.000)	(0.287)	(0.018)	(0.205)
Rank	-0.131***	-2.143***	-0.00003**	-0.00004**	-0.330***	-0.004***	-
	(0.007)	(0.151)	(0.000)	(0.000)	(0.033)	(0.001)	-
Observations	50,278	50,278	50,278	50,278	50,278	40,239	50,278
R-squared	0.347	0.411	0.416	0.129	0.185	0.220	0.509
Control Mean	72.33	1546.18	6.47	0.06	78.62	93.69	130.52
Locality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Qualification FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Specialization FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Cluster Robust Standard errors are in parentheses. Standard errors are clustered at the practice locality level to account for the local labour markets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Panel A shows the coefficient on the Female dummy from estimating equation (1) with only the fixed effects.Panel B shows the coefficients from estimating equation (1). Control mean refers to the average outcome for male doctors across all columns

Table 4: Interaction Effects with Rank from the Discrete Choice Experiment

	Rank on Platform		Respondent demographics		Perceptions about Doctors		
Z	Rank	First Rank	Female	Masters/PhD	Prefer Male	Male Experienced	Male Skilled
Female× Z	0.0036	0.0393	0.0412**	0.0289**	-0.0321***	-0.0092	0.0021
	(0.0252)	(0.0465)	(0.0164)	(0.0124)	(0.0070)	(0.0092)	(0.0104)
consultation_Fee× Z	0.0001	-0.0001	-0.0000	-0.0000	0.0001***	-0.0000	-0.0000
	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
experience× Z	-0.0024	0.0012	0.0001	0.0015***	0.0008**	0.0022***	0.0018***
	(0.0015)	(0.0037)	(0.0005)	(0.0004)	(0.0003)	(0.0003)	(0.0003)
rating× Z	0.0164	-0.0724**	-0.0006	0.0253***	0.0079**	0.0048	0.0002
	(0.0178)	(0.0297)	(0.0048)	(0.0041)	(0.0036)	(0.0038)	(0.0043)
Observations	24,096	24,096	23,916	24,096	23,916	23,916	23,916
R-squared	0.1396	0.1410	0.1362	0.1365	0.1367	0.1372	0.1367

Note: Robust standard errors clustered at the practice locality level are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The reported coefficients are the coefficients of the interaction terms between the variables mentioned in the rows and the corresponding column headers.

Table A.1: Regression Results: Effect of Female Doctor and Controls

VARIABLES	(1)	(2)	(3)	(4)	(5)
Female	-0.0421**	-0.0420**	-0.0419**	-0.0420**	-0.0420**
	(0.0182)	(0.0183)	(0.0183)	(0.0183)	(0.0181)
Consultation Fee	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Experience	0.0057***	0.0059***	0.0059***	0.0055***	0.0055***
	(0.0008)	(0.0009)	(0.0009)	(0.0010)	(0.0009)
Rating	0.1333***	0.1333***	0.1332***	0.1301***	0.1338***
	(0.0080)	(0.0083)	(0.0083)	(0.0082)	(0.0082)
Observations	24,096	24,096	24,096	24,096	24,096
Choice-set FE	No	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes
Rank FE	No	No	No	Yes	No
First Rank FE	No	No	No	No	Yes

Note:This table presents regression estimates for the effect of doctor gender and other covariates on the outcome of interest. Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.2: Sentiment Scores on Text Reviews

Specialization	**Gender**	**Compound Sentiment Score**	**Positive Sentiment Score**
		(1)	(2)
Cardiologists	Male	0.693	0.380
	Female	0.611	0.379
Dentists	Male	0.686	0.410
	Female	0.714	0.418
General Physicians	Male	0.586	0.353
	Female	0.592	0.361
Paediatricians	Male	0.646	0.379
	Female	0.638	0.375
Gynaecologists	Male	0.794	0.414
	Female	0.769	0.395
Orthopaedics	Male	0.535	0.344
	Female	0.407	0.250

Figure A.1: Relative Frequency of Positive Sentiment Words in Text Reviews for Male and Female Physicians