

Brahma, Dweepobotee; Sangwan, Nikita

Working Paper

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

QBS Working Paper, No. 2026/02

Provided in Cooperation with:

Queen's University Belfast, Queen's Business School

Suggested Citation: Brahma, Dweepobotee; Sangwan, Nikita (2026) : The Un'Healthy' Gaps: Evidence on Gendered Faultlines in Digital Healthcare Services, QBS Working Paper, No. 2026/02, Queen's University Belfast, Queen's Business School, Belfast

This Version is available at:

<https://hdl.handle.net/10419/336962>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



**QUEEN'S
UNIVERSITY
BELFAST**

**QUEEN'S
BUSINESS
SCHOOL**

Working Paper Series - QBS Working Paper 2026/02

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

Dweepobotee Brahma

Indian Institute of Technology Jodhpur

Nikita Sangwan

Queen's University Belfast

16 February 2026

Series edited by Philip T. Fliers.

To submit forward your paper to qbs.rps@qub.ac.uk.

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

Dweepobotee Brahma[†] Nikita Sangwan[‡]

February 16, 2026

Abstract

We study whether digital platforms for high-skilled professionals level the playing field or reproduce traditional gender inequalities. Using high-frequency data on physicians, we examine gender differences in labor supply, pricing, patient engagement, and platform visibility. Although the platform equalizes supply-side margins of working hours and fees, female physicians experience lower demand, reduced search visibility, and lower reputation metrics. Investigating the underlying mechanisms, experimental evidence indicates taste-based discrimination, while text-analysis of patient reviews finds no gender differential. These findings underscore the potential role of platforms in reducing institutional constraints but not demand-side biases, with reputation metrics playing a crucial mitigation role.

JEL classification: J16, J24, I11, O33, L86

Keywords: gender bias, digital platforms, healthcare, high-skilled professionals

*We thank Farzana Afridi and Siwan Anderson for constructive feedback. We also thank the attendees at the Queen’s University Belfast Seminar, Global Development Institute (Manchester) Seminar, Liverpool Management School Seminar, IHEA and ACEGD Conference for helpful comments. We acknowledge excellent research assistance from Khushal Sindhav, Abhijeet Balaji Chavan and Cara Taggart. The authors acknowledge the financial support from the Google India Research Award 2021 and Professional Development Award (IIT Jodhpur). This study is registered at AEA RCT Registry ID AEARCTR-0016830. The usual disclaimers apply.

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

[‡]Queen’s Business School. Queen’s University Belfast. Email: n.sagwan24@qub.ac.uk

1 Introduction

Digital platforms are increasingly mediating labor markets across a wide range of sectors, raising fundamental questions about how algorithms and user behavior shape economic opportunities. These platforms are often viewed as equalizing institutions that have the potential to reduce frictions observed in the traditional labor markets by standardizing access, lowering entry barriers, and limiting the scope for discrimination through algorithmic matching and transparent pricing (Horton, 2017). Healthcare has emerged as a particularly important domain of platformisation, especially in low- and middle-income countries where digital health platforms are expanding access to medical services and reshaping how providers and patients interact. These digital transitions are happening at a time when gender inequality remains a persistent feature of healthcare labor markets. Female practitioners often face lower earnings, fewer opportunities, and slower career progression, despite comparable qualifications (Jena *et al.* (2016a), Goldin & Katz (2016)).

This paper studies whether digital platforms mitigate or transmit gender inequality in high-skilled service provision. We examine gender disparities among physicians on a large digital health platform in India, a context where digital healthcare platforms have rapidly expanded access. Unlike traditional labor markets, platform-mediated healthcare markets offer flexible work arrangements, short-duration engagements, transparent pricing, and review systems. These features standardise access to information and lower barriers to participation. However, these features may also introduce new sources of inequality. Algorithmic ranking systems, biased user preferences, and thereby their reputational feedback can embed and amplify inherent biases, potentially reproducing traditional gender gaps even in settings where objective performance and pricing are observable (Barach & Horton (2021), Cowgill & Perkowski (2024)). While a growing literature studies gender inequality on digital labor platforms, evidence from high-skilled, professional service platforms, particularly in healthcare, remains scarce.

This paper contributes to filling this gap by examining labor market outcomes for healthcare providers operating on a large-scale digital health platform. Using novel high-frequency data on over 7,000 physicians between September and December 2024, we document three main findings. First, male and female physicians exhibit statistically indistinguishable labor supply and pricing behavior. Second, despite this supply-side parity, female physicians receive significantly fewer bookings, appear lower in search rankings, and obtain fewer ratings and recommendations. Third, these gaps persist across specifications and subsamples, mirroring patterns commonly observed in offline labor markets for high-skilled professionals (Blau & Kahn, 2017).

A key challenge in interpreting these disparities is distinguishing between statistical discrimination, i.e., when users rationally infer unobserved quality from gender and taste-based discrimination driven by user preferences. This challenge is further amplified by interaction with algorithmic bias. It is plausible that the lower booking rates for female doctors are arising from a reduced visibility due to lower rankings, from fewer or less informative patient ratings and recommendations. To disentangle these mechanisms, we combine observational platform data with two complementary approaches. First, we conduct a laboratory experiment using a simulated digital health platform with fictitious doctor profiles generated using generative AI. The experimental design randomizes physician gender while holding all other attributes constant. The experimental results provide direct evidence of taste-based discrimination, with users exhibiting lower willingness to engage with female physicians even in the absence of observable credentials and quality differences. Second, we analyze the text of patient reviews using natural language processing techniques. We find no systematic gender differences in review content or sentiment, suggesting that perceived quality and patient experiences do not differ by physician gender. It’s important to note that these reviews are left by verified patients who to start with might have been less biased to choose a female physician.

These findings have two important implications. On the one hand, they show that digital platforms can successfully neutralize institutional and supply-side sources of inequality such as differential access, pricing power, or working hours, thereby creating a level playing field in formal terms. On the other hand, they demonstrate that platforms remain vulnerable to demand-side biases, which can reproduce traditional labor market inequalities even in technologically mediated environments.

At the same time, the absence of gendered differences in review content points to a potential corrective mechanism. Reviews provide repeated, public, and information-rich signals of provider quality, which may help users update biased priors over time (Tadelis, 2016). Our results therefore suggest that while digital platforms do not automatically eliminate discrimination, well-designed reputation systems may attenuate its effects as information accumulates over time. As the volume of reviews grows and experience accumulates, these informational signals could reduce reliance on gender-based heuristics and attenuate taste-based discrimination.

This paper speaks to several streams of the literature. First, while this paper relates to the larger literature on gender gaps in the labor market, it specifically contributes evidence from a high-skilled, digitally mediated setting in the Global South, where evidence remains limited. 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 is also connected to the stream of research investigating gender gaps in physician outcomes. Gender gaps in earnings in traditional healthcare markets are attributed either to lower labor supply ([Kehrer \(1976\)](#), [Ohsfeldt & Culler \(1986\)](#), [Sasser \(2005\)](#)), productivity ([Langwell, 1982](#)), marriage or children ([Sasser, 2005](#)), discrimination ([Ohsfeldt & Culler \(1986\)](#), [Chen \(2024\)](#)), differential interpretation of quality signals ([Sarsons, 2017](#)). Beyond earnings, gender gap also exist through slower career progression [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.* \(2016b\)](#) [Lo Sasso *et al.* \(2020\)](#), [Catenaccio *et al.* \(2022\)](#)). [Chen \(2024\)](#) is the only study to our knowledge that investigates gender gaps in the digital health space in developing economy of China. This is the first paper to examine gender disparities on a digital health platform in India.

This paper makes several contributions to the literature on gender inequality in platform-mediated labor markets. First, we construct a novel and highly granular dataset comprising over 50,000 observations on more than 7,000 physicians from a leading digital health platform in India. The richness of the platform data allows us to observe physicians’ labor supply decisions along both the extensive and intensive margins, alongside detailed measures of engagement, pricing, and visibility.

Second, the paper adopts a multi-method approach that combines observational platform data, a laboratory experiment, and textual analysis of patient reviews using natural language processing techniques. This integrated design enables us to build a comprehensive body of evidence on both the magnitude and the mechanisms underlying gender gaps among high-skilled workers on digital platforms. The laboratory experiment allows us to isolate the role of users’ prior beliefs and preferences as a source of bias, distinguishing taste-based discrimination from platform-driven mechanisms. Complementing this, the NLP-based analysis of patient reviews reveals no systematic gender differences in review sentiment or content, thereby ruling out statistical discrimination operating through reputational feedback. To our knowledge, this represents one of the first applications of NLP methods in health economics to study gender bias in patient-generated text data.

Finally, this study contributes new evidence from a developing-country context. While existing research on gender gaps in digital labor markets has largely focused on advanced economies and lower-skilled platforms, our analysis of a digital health platform in India

extends the literature to high-skilled professional services in the Global South. To the best of our knowledge, this is the first study to examine gender disparities on a digital health platform in India.

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 discusses the total evidence across the three approaches and section 7 concludes with a discussion on limitations and future directions.

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 an e-commerce website with a product search option. 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 reviews¹. 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 “recommendation score”, defined as the percentage of all recommendations that are positive.²

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

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

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.

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, column (3) reports the average for female doctors and column (4) reports the raw gender

differences in these variables (i.e., male – female averages). 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 about 21 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 compared to the female doctors’ average of 1714.63 minutes per week. Female doctors offer slightly longer slots than male doctors on average. Despite female physicians charging a lower average consultation fee (696 INR) compared to male doctors (780 INR), they have a lower booking rate. This result is in line with the findings of [Chen \(2024\)](#) for China. On the platform, female physicians have fewer reviews left by patients. Interestingly, female doctors have a marginally higher average recommendation rate. However, female doctors are ranked about 60 lower ranks on average compared to their male counterparts.

Our sample spans six specialties that are 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 specialties, male doctors outnumber female doctors. Digital health platforms are largely prevalent in urban areas with very little penetration in 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 remaining share of 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, and asked to choose a doctor from each set. The attributes included physician gender (male

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

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:

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

where y_{icts} is the outcome variable for doctor i in locality 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 locality 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 doctor 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., doctors), 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 doctor 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 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

5.1 Platform data

Table 3 presents the regression results based on the platform data. The estimates include fixed effects for doctors’ geographical locality of practice, the week in which the data were scraped, as well as doctors’ qualification and specialization. Standard errors are clustered at the locality level, as doctors primarily serve local markets, and unobserved factors affecting their outcomes may be correlated within a locality. This specification allows us to isolate the variation arising from local labor market competition. Even after accounting for these fixed effects some of the significant gender gaps observed in the raw gender gaps in the results between male and female physicians (refer Table 1) continue to persist.

Panel A of Table 3 reports the estimates including only the fixed effects. We find no statistically significant difference between male and female doctors in their availability on the platform, measured either on the extensive margin (number of available slots; column 1) or the intensive margin (total duration of availability; column 2). Similarly, there is no significant gender difference in the (log) consultation fees charged (column 3). Hence, the raw average gender differences in platform services disappear once we account for local market conditions.

However, despite the lack of gender differences in service availability or pricing, there is a pronounced gender gap in the demand for these services. Booking rates for female doctors are lower by 1.4 percentage points (23.33%, $p < 0.01$) relative to male doctors (column 4). Female physicians also receive 59.78% ($p < 0.01$) fewer patient reviews (column 5). By contrast, there is no significant gender difference in the likelihood that patients recommend doctors to others (column 6). Finally, female doctors are 13.5% ($p < 0.01$) more likely to be ranked lower on the platform compared to their male counterparts (column 7).

Panel B of Table 3 extends the specification in the top panel by incorporating additional controls that may influence patients’ choice of doctors on the platform. Specifically, we control for whether the doctor practices in a hospital, years of experience, and their rank on the platform. Even after including these doctor-level characteristics, the results remain qualitatively similar to those in Panel A. There continues to be no significant gender difference in doctors’ availability on the platform (columns 1 and 2) or in the consultation fees they charged (column 3). However, female doctors still experience a significantly lower booking rate than their male counterparts, though the magnitude of the gap decreases to 0.011 percentage points (18.33%, $p < 0.01$) once these additional controls are included (column 4).

On the demand side, the gender gap in the number of reviews remains large and statistically significant, with female doctors receiving 44.79% fewer reviews ($p < 0.01$) (column

5). Interestingly, the previously insignificant difference in recommendation rates now becomes marginally significant, suggesting a gender gap of 0.84 percentage points (column 6). Finally, female doctors continue to be more likely to hold a lower rank on the platform by 9.11% ($p < 0.01$) compared to their male counterparts (column 7).

Across both specifications, there are no significant gender differences in doctors' availability or consultation fees, indicating that male and female physicians supply services on the platform at comparable levels and prices. However, female doctors consistently experience lower demand, as reflected in significantly lower booking rates and fewer patient reviews. These gaps persist even after controlling for observable doctor characteristics such as experience, hospital affiliation, and platform rank, though their magnitudes decline somewhat. Notably, our results remain qualitatively similar when we restrict the sample to specializations with a more balanced gender representation and exclude cardiology, orthopaedics, and gynaecology (refer Table A.2).

These findings raise an important question: what explains the lower booking rates for female doctors despite comparable labor supply and pricing? Are these differences primarily driven by disparities in platform ranking, or by consumer-side factors? To isolate the role of these metrics, the next section will discuss the results of a lab experiment designed to test how consumers respond to identical doctor profiles that vary only by gender.

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 from the discrete choice experiment for specific physician attributes on the platform. The choice experiment reveals that consumers exhibit a negative willingness to pay for female doctors (-225.66, 95% CI: -412.10 to -39.22)(column 1), indicating a significant gender bias even after holding other attributes constant. The respondents willingness to pay INR 225 less to consult with a female physician provides direct evidence of a monetary penalty associated with doctor's gender. On expected lines, willingness to pay increases with experience (31.93, CI: 14.20-49.66)(column 2) and especially with higher patient ratings (715.78, CI: 339.13-1092.44)(column 3). These results suggest that while participants value quality indicators, they systematically discount

female providers, highlighting the role of taste-based discrimination in driving lower demand for female doctors.

The baseline regression results of the discrete choice experiment are presented in Panel A of Table 4. These reveal systematic gender differences in consumer preferences when selecting doctors. Across all specifications, the coefficient on the female indicator remains negative (approximately -0.042) and statistically significant ($p < 0.05$), indicating that consumers are less likely to choose a female doctor compared to an otherwise identical male doctor. This gender gap persists even after accounting for choice-set, individual, and rank fixed effects, suggesting that the bias is not explained by unobserved heterogeneity in respondents or contextual factors within the choice environment. Consistent with standard demand patterns, higher consultation fees reduce the likelihood of selection, while greater experience and higher patient ratings significantly increase it. Overall, these results highlight that although consumers respond rationally to price and quality signals, they systematically penalize female doctors, providing causal evidence of consumer-side discrimination.

The interaction regressions in Panel B of Table 4 further investigate whether the observed gender bias varies with doctors' rank and respondents' perceptions about doctors' gender and competence. The results show that the negative effect of being female on doctor choice does not significantly differ by platform rank (columns 1-2), suggesting that ranking visibility does not drive respondent choices.

We also leverage respondents' stated preferences to examine whether these mediate choices by interacting with physician and respondent attributes. Notably, the interaction between the female indicator and respondents who *prefer male doctors* is negative and highly significant (-0.0321, $p < 0.01$) (column 3), indicating that gender bias is particularly pronounced among consumers who explicitly state a preference for male providers. The coefficients on other perception-based interactions such as viewing male doctors as more experienced or skilled are small and statistically insignificant (columns 4 and 5).

Overall, these findings suggest that taste-based discrimination arises primarily from explicit gender preferences rather than from perceived differences in ability or platform visibility. These findings provide evidence of taste-based discrimination rooted in prior beliefs.

5.3 Textual Analysis

Having ruled out physician rank as the primary mechanism behind the 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 to investigate this further through the platform data. Since healthcare is a credence good, reputation metrics such as patient reviews and

recommendations play a crucial role in shaping consumer choices. Thus, systematic gender differences in how patients evaluate and review doctors could further reinforce unequal demand. If a potential patient trying to choose a doctor on the platform finds that female doctors are on average rated lower than male doctors, and chooses to make an appointment with a male doctor, that would be indicative of statistical discrimination. To investigate this possibility, we analyse the text and sentiment of patient reviews, which underpin physicians’ recommendation scores on the platform, to assess whether female doctors receive systematically lower feedback in tone or content compared to their male counterparts.

To do so, we scraped a random sample of 165,826 text reviews from the sample of doctors analyzed through the platform data⁴ from the platform. Of these, 82,734 reviews correspond to male doctors and 83,091 to female doctors. Since multiple reviews can be associated with each physician, this dataset spans 4,464 unique doctors, comprising 2,209 male and 2,255 female physicians. All the analysis using this text data are normalized using the review count so that doctors with more number of reviews do not dominate the results.

Reviews for male and female doctors are broadly similar. Both receive nearly identical average tag counts, though male doctors are slightly more likely to receive all five available tags, and each tag appears with comparable frequency across genders.⁵ Reviews for female doctors are somewhat longer, averaging 43.9 words versus 33.6 for male doctors.⁶ Analysis of parts of speech -adjectives, adverbs, intensifiers, and superlatives - shows no significant gender differences in usage.⁷ We use measures of *polarity* and *subjectivity* to understand the sentiment of the reviews. Polarity refers to the overall sentiment of the text, with -1 denoting extreme negative and $+1$ denoting extreme positive. Subjectivity quantifies the amount of personal opinion or factual information contained in the text, with a score of $+1$ implying a text is fully subjective and a score of 0 implying that the text is fully objective.⁸

We use `TextBlob` which employs a lexicon-based approach that tokenizes review text and compares it to a predefined dictionary classifying words as positive or negative, assigning each a polarity and subjectivity score. These are aggregated to produce a composite score for each review. The left panel of Figure 1a shows the distribution of polarity scores, and the right panel shows subjectivity scores for reviews of male and female doctors. Polarity

⁴For computational ease, we scrape a random sample of reviews belonging to each of the doctors in our sample from the platform data. For doctors who had 200 reviews or less we tried to scrape all the reviews. For doctors with larger number of reviews, we scrape an average of 60% random sample of their text reviews.

⁵refer Appendix Figure A.1.

⁶refer Appendix Figure A.2.

⁷refer Appendix Figure A.2b and Figure A.2c.

⁸An example of a completely subjective review would be “*I felt completely cared for and truly believe Dr. Smith is the most compassionate and understanding physician I’ve ever met*”, while an example of a completely objective factual review would be “*During my visit, Dr. Smith conducted a physical examination, ordered blood tests, and provided a prescription for treatment.*”.

distributions are similar across genders, indicating that reviews for both male and female doctors are generally positive, consistent with overall recommendation scores in the nineties.⁹ The right panel in Figure 1a shows a spike at 0 for reviews for male doctors than female doctors implying that reviews for male doctors tend to be more objective than those for female doctors.

We analyze the actual language used in text reviews to examine gender-based differences in how doctors are described. To guide this analysis, we draw on the Stereotype Content Model (Fiske *et al.* (2002); Cuddy *et al.* (2007)), which identifies two core dimensions of social perception: warmth and competence. Using Empath (Fast *et al.* (2016)) - a pretrained library with numerous lexicon categories - we select categories aligned with these two dimensions. Each category begins with a set of seed words, which are then expanded into larger word corpora through word embeddings that capture semantically similar terms. For example, the category communication includes related terms such as language and signals. We then measure, by gender, the frequency of words falling into each lexicon category.

Panels B and C of Figure 1 present these results. For warmth-related categories (Panel B, Figure 1b), there is no statistically significant difference in the average score across reviews for male and female doctors. The only difference appears in the “politeness” and “affection” categories where female doctors appear to have a higher average score than male doctors. In case of competence-related categories (Panel C, Figure 1c), generic professional topics such as ‘office’ and ‘communication’ are used equally across genders and show no statistical difference. Discussions directly tied to medical expertise - such as those in the categories of ‘health’, ‘healing’, and ‘medical emergency’ - are more common in reviews of male doctors than female doctors. This suggests that explicit references to medical competence are more likely to be present in reviews of male doctors than for female doctors. However, considering the totality of evidence across different metrics from the text reviews, we don’t find strong evidences of systematic gender differences in patient feedback.

6 Discussion

Considering the evidence across the three methods provides several important insights. Analysis of the platform data shows that, after controlling for observable characteristics and local labor market conditions, male and female doctors have no difference in labor supply or prices charged. Despite this, female doctors are chosen less frequently, as reflected in lower booking rates, and consistently appear lower in the platform’s search results. They

⁹A complementary sentiment analysis using NRCLex Mohammad & Turney (2013), shown in Figure A.2c in the Appendix, confirms these results.

also receive lower reputation scores, with fewer patient reviews and lower recommendation ratings compared to male counterparts. Since the lower booking rates could mechanically contribute to fewer reviews and recommendations, it is important to account for this scale effect. Conditioning on the total number of patients seen on the platform, the gender gap in the count of reviews disappears, while the gap in recommendation ratings narrows to 0.6 but remains statistically significant (refer Appendix Table [A.1](#)). This suggests that, even after accounting for patient volume, female doctors receive similar numbers of reviews but slightly lower recommendation ratings, pointing to persistent gender-based bias in patient evaluations.

The laboratory experiment was designed to disentangle the effects of rank and recommendation on the gender gap in booking rates. The simulated platform is designed to display two doctors (one male and one female) at the same rank with other characteristics built using information from the platform data systematically varied across the choice sets. This allows us to hold rank constant and systematically vary other characteristics holding them constant one at a time. We find a negative willingness to pay for female doctors even when displayed at the same rank as a male doctor, and interaction of rank with gender showed no gender gap in the probability of being chosen by the respondent. Thus the finding of lower probability of a female doctor being chosen relative to the male doctor despite identical characteristics is indicative of taste-based discrimination. This is further strengthened by the consistent negative coefficient on the female dummy in the within-respondent specification and the negative coefficient in the interaction between the female dummy and prior preferences over male doctors. While the patients on the digital platform are not comparable to the undergraduate student respondents from the laboratory experiment, the experiment serves an useful additional purpose. It allows us to rule out rank on the platform as a source of statistical discrimination in addition to other observable characteristics. Natural Language Analysis of the text data from patient reviews provides further evidence of taste-based discrimination over statistical discrimination. Barring small linguistic differences in the content of reviews, there are no systematic differences between the reviews of male and female doctors across multiple metrics of tone, sentiments or language usage. This finding rules out the various reputation metrics as the source of statistical discrimination.

Having ruled out statistical discrimination as the reason behind the observed lower booking rates for female doctors on the platform, we can conclude that it is plausibly driven by taste-based discrimination stemming from the patients implicit biases.

7 Conclusion

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 on doctors from a widely used telemedicine 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 compared to their male counterparts. Observational data alone cannot disentangle the mechanisms driving lower booking rates - whether these stem primarily from reduced visibility due to lower ranks, or from bias stemming from prior patient beliefs. 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, textual analysis of patient reviews indicates that reviews are largely positive for both male and female doctors and there are no significant gender differences in how patients review male and female doctors. The totality of evidence across observational data from the platform, laboratory experiment and textual analysis enables us to rule out all potential channels through which statistical discrimination could be the driver of the observed lower booking rates for female doctors relative to male doctors.

While this is the first study on the private practice market of doctors in India, it has some limitations. First, digital health platforms in India remain largely concentrated in urban areas, and thus this data covers the six major cities in India and may not generalise to rural settings. Second, the platform serves both as a booking interface and a directory, and we cannot observe consultations that are arranged off-platform after search. Unless such off-platform behaviour differs systematically by physician gender, this limitation is unlikely to bias our estimates. Third, the platform does not provide information on the gender of reviewers and allows anonymous reviews, preventing a deeper examination of whether reviewer gender interacts with review language.

Despite these limitations, the findings offer several important insights. First, in contrast to existing evidence of gender gaps in physicians' labor supply, we find no such differences in this platform setting, suggesting that the flexibility offered by digital platforms may reduce supply-side gender constraints. This is particularly relevant in the Indian context, where women constitute a substantial share (and in some specializations, a majority) of medical graduates. At the same time, our results demonstrate that equal participation and equivalent

performance do not guarantee equal outcomes. Even in a high-skilled professional market with transparent pricing and observable quality signals, user biases can persist and translate into differential demand.

More broadly, this study contributes to a growing literature on how digital platforms and algorithmic systems shape inequality in the digital economy. By focusing on a high-skilled healthcare market in a developing-country context, it highlights a potential role for reputational mechanisms in mitigating the traditional gender biases over time. These findings underscore the importance of platform design in amplifying high-quality, informative feedback and suggest that well-structured reputation systems can serve as a countervailing force against bias in digital labor markets.

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.
- Barach, Moshe A, & Horton, John J. 2021. How do employers use compensation history? Evidence from a field experiment. *Journal of Labor Economics*, **39**(1), 193–218.
- Bashaw, David J., & Heywood, John S. 2001. The Gender Earnings Gap for US Physicians: Has Equality been Achieved? *Labour*, **15**(3), 371–391.
- Blau, Francine D, & Kahn, Lawrence M. 2017. The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, **55**(3), 789–865.
- 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.
- Cowgill, Bo, & Perkowski, Patryk. 2024. Delegation in hiring: Evidence from a two-sided audit. *Journal of Political Economy Microeconomics*, **2**(4), 852–882.
- Cuddy, A. J. C., Fiske, S. T., & Glick, P. 2007. The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, **94**(4), 631 – 648.
- Fast, Ethan, Chen, Binbin, & Bernstein, Michael S. 2016. Empath: Understanding Topic Signals in Large-Scale Text. *Pages 4647 – 4657 of: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI'16. New York, NY, USA: Association for Computing Machinery.

- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, **82**(6), 878 – 902.
- 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. *Proceedings of the ACM on Human-Computer Interaction*, **2**(CSCW), 1–21.
- Goldin, Claudia, & Katz, Lawrence F. 2016. A most egalitarian profession: pharmacy and the evolution of a family-friendly occupation. *Journal of Labor Economics*, **34**(3), 705–746.
- 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.
- Horton, John J. 2017. The effects of algorithmic labor market recommendations: Evidence from a field experiment. *Journal of Labor Economics*, **35**(2), 345–385.
- Jena, Anupam B, Olenski, Andrew R, & Blumenthal, Daniel M. 2016a. Sex differences in physician salary in US public medical schools. *JAMA internal medicine*, **176**(9), 1294–1304.
- Jena, Anupam B., Olenski, Andrew R., & Blumenthal, Daniel M. 2016b. 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.
- Mohammad, Saif, & Turney, Peter. 2013. Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*, **29**(3), 436 – 465.
- 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.
- Tadelis, Steven. 2016. Reputation and feedback systems in online platform markets. *Annual review of economics*, **8**(1), 321–340.
- Theurl, Engelbert, & Winner, Hannes. 2011. The male - female 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 (type of practice)	0.42	0.46	0.38	0.08***
Available Slots (count)	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 (INR)	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 (count)	71.30	78.62	64.15	14.47***
<i>Qualifications</i>				
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	0.08***
Dentist	0.37	0.33	0.41	-0.08***
General Physician	0.15	0.20	0.10	0.10***
Gynaecologist	0.21	0.03	0.39	-0.36***
Orthopaedist	0.11	0.22	0.00	0.22***
Paediatrician	0.11	0.12	0.10	0.02***
Panel C: City Composition of the Physicians in our Sample				
Bangalore	0.35	0.32	0.38	-0.06***
Chennai	0.16	0.18	0.14	0.04***
Delhi	0.16	0.16	0.16	0.00***
Hyderabad	0.15	0.14	0.15	-0.01***
Kolkata	0.06	0.08	0.04	0.04***
Mumbai	0.13	0.12	0.14	-0.02***

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 hospital and zero otherwise. Available slots are the total number of slots offered over a week at the time of scraping the data. Available duration is the total minutes over this week. Booking rate is calculated as the percentage of total booked slots of the total available slots for the upcoming week at the time of scraping the data. The *Qualifications* indicate the highest degree attained of the physician. Column (4) in each panel reports the difference between male and female physicians. 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
	(1)	(2)	(3)
Age	265	20.079	2.815
Female respondent	269	0.130	0.337
Male respondent	269	0.862	0.345
Holds Bachelor's degree	269	0.807	0.396
Holds Master's/PhD	269	0.193	0.396
Exposure to doctors			
<i>Familiar with doctor</i>	267	2.494	1.249
<i>Family doctor</i>	269	0.335	0.473
<i>Father is a doctor</i>	267	0.030	0.171
<i>Mother is a doctor</i>	267	0.022	0.148
<i>Brother is a doctor</i>	267	0.075	0.264
<i>Sister is a doctor</i>	267	0.060	0.238
<i>Other male member is doctor</i>	267	0.202	0.402
<i>Other female member is doctor</i>	267	0.082	0.275
<i>Last consulted doctor female</i>	269	0.275	0.447
<i>Last consulted doctor male</i>	269	0.710	0.455
Perception about doctors			
<i>Male doctor perceived more experienced</i>	267	0.494	0.501
<i>Male doctor perceived more skilled</i>	267	0.408	0.492
<i>Preference for male doctor</i>	267	0.404	0.492
Panel B: Estimates of Willingness to Pay from the Choice Experiment			
	Female	Experience	Rating
	(1)	(2)	(3)
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

Note: Panel A summarises the demographic characteristics, prior exposure to doctors and gender perceptions about doctors held by the 269 respondents in the lab experiment. Panel B reports the willingness-to-pay (WTP) estimates and the Lower and upper limit on these estimated from the discrete choice experiment for the specific physician attribute of gender (Female in column (1)), years of experience (Experience in column (2)) and rating (column (3)).

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 (1.752)	25.439 (33.287)	-0.010 (0.024)	-0.014*** (0.003)	-47.006*** (5.394)	-0.144 (0.333)	17.581*** (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 (1.676)	4.038 (32.185)	0.031 (0.025)	-0.011*** (0.004)	-35.216*** (4.605)	-0.841** (0.331)	11.891*** (3.513)
Hospital	-8.442** (3.342)	-145.836** (68.184)	0.180*** (0.045)	0.017*** (0.006)	-39.834*** (6.967)	-2.254*** (0.563)	8.923 (7.493)
Experience	-0.470*** (0.099)	-16.873*** (1.875)	0.010*** (0.001)	0.000 (0.000)	2.220*** (0.287)	-0.225*** (0.018)	-1.842*** (0.205)
Rank	-0.131*** (0.007)	-2.143*** (0.151)	-0.00003** (0.000)	-0.00004** (0.000)	-0.330*** (0.033)	-0.004*** (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
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: Panel A reports the regression coefficient on the Female dummy from estimating equation (1) with fixed effects. Panel B incorporates additional controls - *Hospital* that takes a value of one if the type of practice is a hospital, *Experience* is the number of years of experience of the doctor and *Rank* at which the profile is displayed on the platform. Control mean reports the mean outcome for the male doctors for the corresponding column. Standard errors are clustered at the practice locality level to account for the local labour markets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

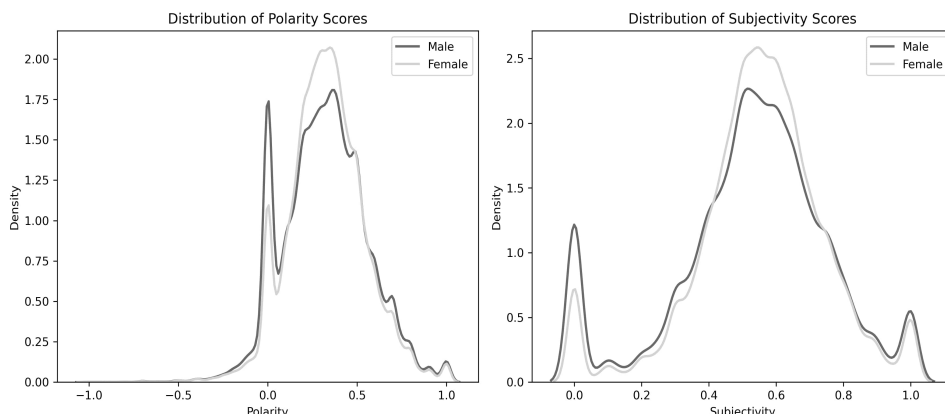
Table 4: Regression Results from the Discrete Choice Experiment

Panel A: Regression Results of Choice of Doctors by characteristics					
	(1)	(2)	(3)	(4)	(5)
Female	-0.0421** (0.0182)	-0.0420** (0.0183)	-0.0419** (0.0183)	-0.0420** (0.0183)	-0.0420** (0.0181)
Consultation Fee	-0.0002*** (0.0000)	-0.0002*** (0.0000)	-0.0002*** (0.0000)	-0.0002*** (0.0000)	-0.0002*** (0.0000)
Experience	0.0057*** (0.0008)	0.0059*** (0.0009)	0.0059*** (0.0009)	0.0055*** (0.0010)	0.0055*** (0.0009)
Rating	0.1333*** (0.0080)	0.1333*** (0.0083)	0.1332*** (0.0083)	0.1301*** (0.0082)	0.1338*** (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

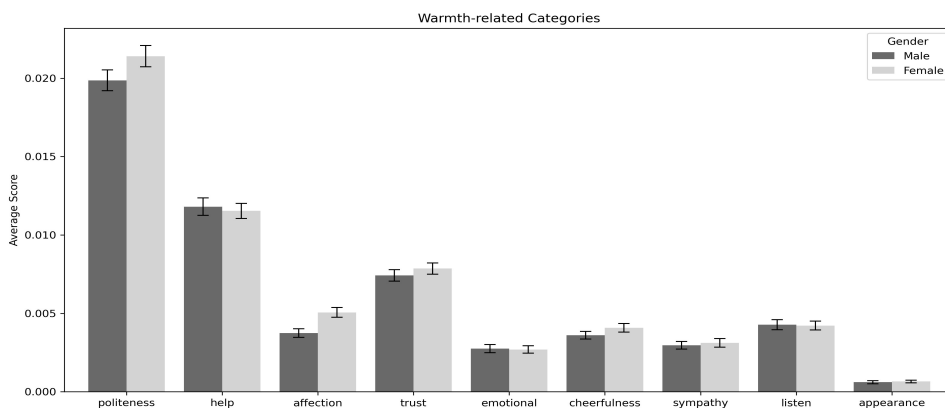
Panel B: Heterogeneity Regression Results by Doctor's Rank and Respondent Perceptions					
Z	Rank on Platform		Perceptions about Doctors		
	Rank (1)	First Rank (2)	Prefer Male (3)	Male Experienced (4)	Male Skilled (5)
Female × Z	0.0036 (0.0252)	0.0393 (0.0465)	-0.0321*** (0.0070)	-0.0092 (0.0092)	0.0021 (0.0104)
Consultation Fee × Z	0.0001 (0.0001)	-0.0001 (0.0001)	0.0001*** (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)
Experience × Z	-0.0024 (0.0015)	0.0012 (0.0037)	0.0008** (0.0003)	0.0022*** (0.0003)	0.0018*** (0.0003)
Rating × Z	0.0164 (0.0178)	-0.0724** (0.0297)	0.0079** (0.0036)	0.0048 (0.0038)	0.0002 (0.0043)
Observations	24,096	24,096	23,916	23,916	23,916
R-squared	0.1396	0.1410	0.1367	0.1372	0.1367

Note: Panel A reports the regression coefficients for the female dummy and other characteristics (consultation fee (INR), Experience (years) and Rating that were randomised across the doctor profiles on the simulated platform. Panel B reports the heterogeneity coefficients of the interaction of the observed characteristics with Rank (column 1), indicator for being on the *First Rank* (column 2), indicator for preference of male doctors (column 3), indicator for the perception of male doctors being more experienced (column 4), and indicator for the perception of male doctors being more skilled (column 5). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

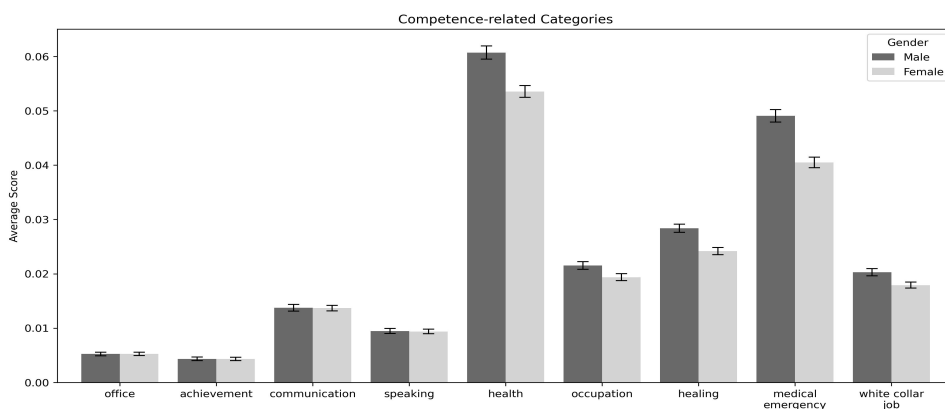
Figure 1: Text Analysis on the Patient Reviews



(a) Panel A: Polarity versus Subjectivity Scores of Reviews by Gender



(b) Panel B: Mentions of Warmth Related Lexicon Categories by Gender



(c) Panel C: Mentions of Competence Related Lexicon Categories by Gender

Note: Panel A illustrates the kernel density distributions of the polarity (left) and subjectivity (right) scores per review across male and female doctors, respectively. Panel B illustrates the average score per lexicon category within the overall 'Warmth' category for male and female doctors normalized by word count and review count. Panel C illustrates the average score per lexicon category within the overall 'Competence' category for male and female doctors normalized by word count and review count. Bars in Panels B and C display 95% bootstrapped confidence intervals constructed using 2000 bootstrap iterations.

A APPENDIX

Table A.1: Gender Gaps and Regression Results Controlling for Patient Count

	Reviews	Recommendation
	(1)	(2)
Female	-1.295 (1.324)	-0.647* (0.334)
Hospital	-3.477* (1.898)	-2.053*** (0.553)
Experience	-0.227*** (0.081)	-0.236*** (0.018)
Rank	-0.035*** (0.007)	-0.003*** (0.001)
Patient Count	0.588*** (0.016)	0.003*** (0.000)
Observations	49,921	40,239
R-squared	0.757	0.225
Locality FE	Yes	Yes
Week FE	Yes	Yes
Qualification FE	Yes	Yes
Specialization FE	Yes	Yes

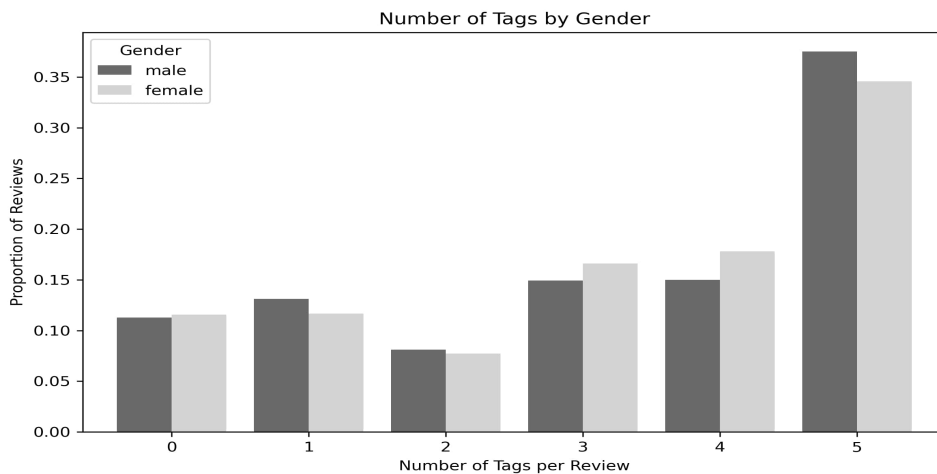
Note: This table reproduces Columns (5)-(6) on *Reviews* and *Recommendation*, respectively, of Panel B in Table 3 with the additional control for patient count. Standard errors are robust and clustered at the practice locality level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.2: Gender Gaps and Regression Results Across Outcomes (for gender-balanced specialisations)

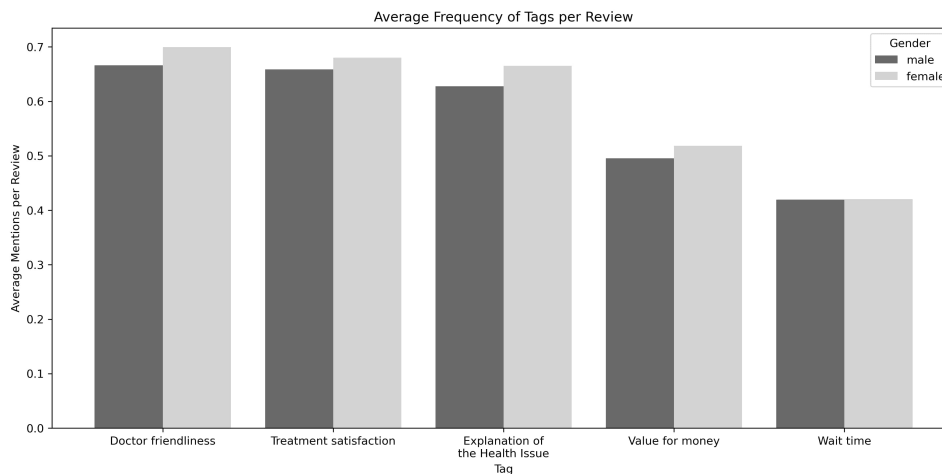
	Available Slots	Available Duration	Log Consultation Fees	Booking Rate	Reviews	Recommendation	Rank
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Regression Results with only Fixed Effects							
Female	-3.454* (1.976)	28.587 (37.600)	-0.044** (0.018)	-0.017*** (0.004)	-56.803*** (5.765)	-0.076 (0.352)	17.451*** (3.795)
Observations	31,736	31,736	31,736	31,736	31,736	24,945	31,736
R-squared	0.334	0.402	0.473	0.175	0.162	0.252	0.494
Panel B: Regression Results with Additional Controls							
Female	-2.943 (1.913)	-6.648 (36.450)	-0.008 (0.016)	-0.014*** (0.004)	-43.397*** (5.013)	-0.584* (0.345)	10.662*** (3.848)
Hospital	-8.564** (4.118)	-163.361* (84.267)	0.236*** (0.034)	0.022** (0.009)	-39.768*** (9.882)	-2.932*** (0.847)	31.606*** (10.085)
Experience	-0.432*** (0.136)	-19.223*** (2.718)	0.008*** (0.001)	0.000 (0.000)	2.333*** (0.349)	-0.147*** (0.023)	-2.122*** (0.275)
Rank	-0.127*** (0.008)	-2.063*** (0.171)	-0.000 (0.000)	-0.000*** (0.000)	-0.347*** (0.035)	-0.003*** (0.001)	- -
Observations	31,736	31,736	31,736	31,736	31,736	24,945	31,736
R-squared	0.385	0.441	0.489	0.179	0.217	0.272	0.506
Control Mean	81.1455	1758.7122	6.3332	0.0755	93.2642	94.6892	162.8715
Locality 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: This table reproduces Table 3 restrict the sample to specializations with a more balanced gender representation, i.e., it includes Dentists, General Physicians and Pediatricians. Standard errors are robust and clustered at the practice locality level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure A.1: Text Analysis on the Patient Reviews

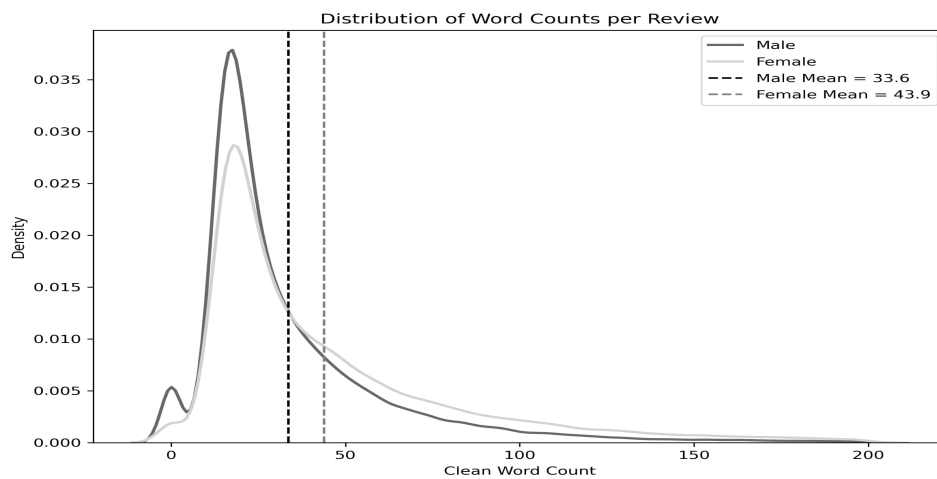


(a) Panel A: Average Counts of Tags per Review by Gender

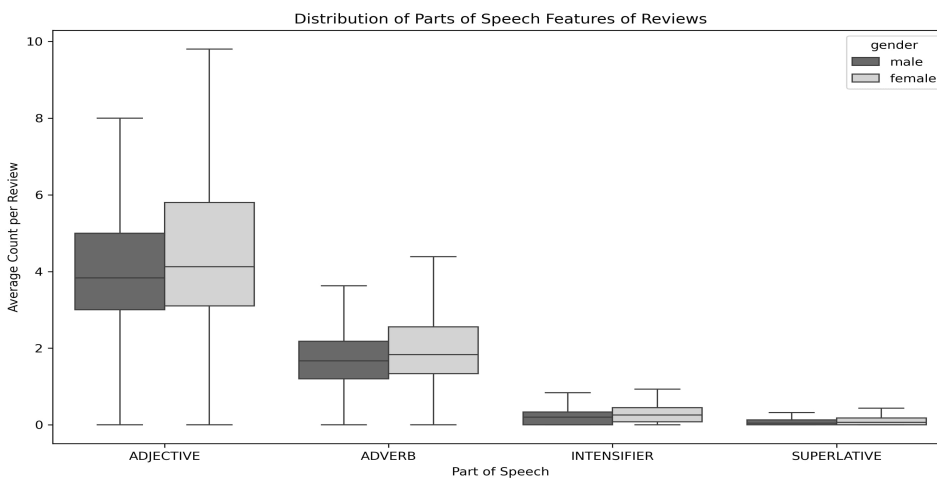


(b) Panel B: Average Counts of Unique Tags per Review by Gender

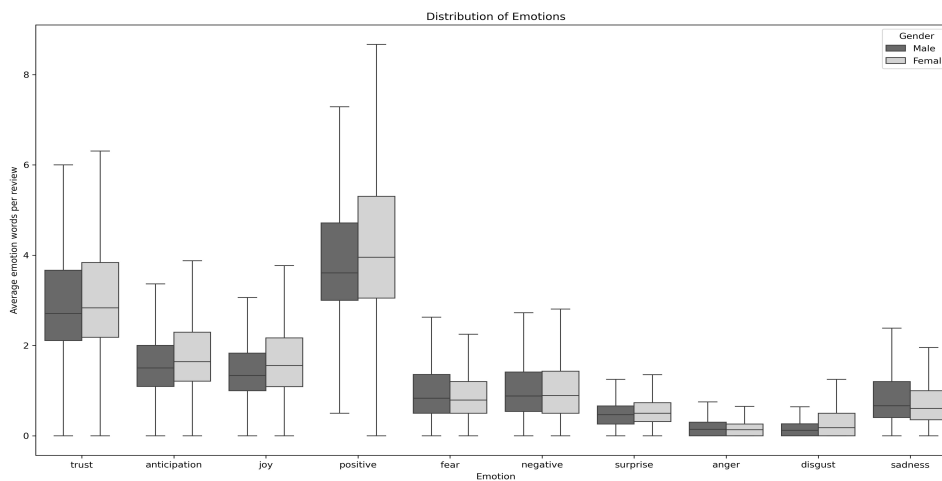
Figure A.2: Text Analysis on the Patient Reviews



(a) Panel A: Word Counts of Reviews by Gender



(b) Panel B: Average Counts of Parts of Speech by Gender



(c) Panel C: Average number of Emotion Words per Review by Gender

Figure A.3: Simulated Experimental Platform

