Worth the Wait? Willingness to Pay for Gender-Concordant Physicians Across Medical Specialties

Helen Kissel,* Tamri Matiashvili,† and Helena Roy‡§

October 6, 2026

Click here for the latest version.

Abstract

Patients increasingly select physicians through online "find a doctor" tools, which display information such as physician gender and patient ratings, rather than relying on physician referral, word-of-mouth, or soonest availability. Medical literature documents female patients' stated preferences for female physicians in certain specialties, but whether and how patients act on these preferences when given the tools to do so remains an open question. Using a randomized hypothetical choice experiment that mirrors the information presented in typical online physician-search platforms, we find that female patients have strong preferences for gender-concordant physicians in gender-sensitive specialties. Women are willing to wait 1.6 additional weeks for a female urologist and 2.2 additional weeks for a female psychiatrist, two specialties with more gender-sensitive components, but show no statistically significant concordance preference in orthopedics, a more gender-neutral specialty. Men exhibit no significant concordance preference in any specialty. Among women, the gender concordance preference is stronger among those who report prior negative healthcare experiences due to gender. We complement these experimental findings with evidence from Medicare claims data on bladder cancer patients. We find that female patients are more likely than male patients to see female urologists, and women who match with a female urologist wait longer than women who match with a male urologist. Male patients exhibit no corresponding wait-time difference by physician gender. These findings suggest that patient preferences over physician characteristics, and patients' growing capacity to act on those preferences, mean that the availability of female healthcare providers may affect patterns of health-seeking behavior such as provider choice and timing of care.

*Kissel: Simon Fraser University – helen_kissel@sfu.ca

†Matiashvili: Stanford University – tamrim@stanford.edu

‡Roy: Hasso Plattner Institute – helena.roy@hpi.de

§We are grateful to Marcella Alsan, Matthew Brown, Liran Einav, Harsh Gupta, Jonathan Lee, Carl Meyer, Petra Persson, Maria Polyakova, Maya Rossin-Slater, Paul Sniderman, Ariel Stern, Michael Tomz, and Heidi Williams for feedback and support. Participants of the Stanford Institute for Theoretical Economics (Economics of Health Inequality) and workshops at Stanford University provided helpful comments. This research was funded by grants from the Institute for Research in the Social Sciences (IRiSS) and the Stanford Institute for Economic Policy Research (SIEPR) at Stanford University. This study was approved by the Stanford University Institutional Review Board (Protocol IDs 84149 and 84962).

1. Introduction

Physician demographic attributes can have profound effects on demand and take-up of medical care (Alsan, Garrick and Graziani, 2019). The medical literature provides several examples of patients expressing a preference for a physician of the same gender, stretching across a range of specialties including gender-sensitive emergency department visits (Nolen et al., 2016), endoscopy procedures (Varadarajulu, Petruff and Ramsey, 2002), urology (McDevitt and Roberts, 2014), and gynecology (Reyes, 2008). These preferences may be consequential if the physician workforce is not representative of the population, and if patients have the ability to act on these preferences. The first condition is known: using national administrative data on roughly 1.19 million physicians, we document that patients' ability to match with a gender-concordant provider varies widely across both specialties and locations. For example, 17% and 10% of Hospital Referral Regions have no female urologists and orthopedists, respectively.

The strength of the expressed gender concordance preferences, and whether they translate into differences in care-seeking behavior and downstream health outcomes, remains largely unknown. This is particularly true following the advent of online "find a doctor" tools, which have enabled patients to review demographic and other physician characteristics online, before scheduling an appointment. Whereas previously patients may have matched with a physician based on a referring physician's recommendation, scheduling considerations, or word-of-mouth, they are now able to review multiple profiles online, often seeing physician photographs and reading other patients' reviews. Thus, patients' ability to act on their preferences has increased in recent years. If patients are willing to delay care to match with a concordant provider, these preferences could have meaningful implications for the timing of treatment and disease progression; something patients may be unable to account for, given the asymmetric information that characterizes healthcare (Arrow, 1963). These preferences may also vary systematically across medical specialties, particularly with the degree of gender sensitivity in care. Understanding whether and how these preferences are acted upon is important for assessing their consequences for both patient health and the allocation of the physician workforce across specialties and geographic areas, where the gender composition of providers may constrain patients' ability to match.

This paper addresses this knowledge gap by using a hypothetical choice experiment to measure the willingness-to-pay (WTP) for gender concordance preferences across different types of specialties, and by documenting how these preferences manifest in real-world patient behavior when access to concordant providers may be limited. The hypothetical choice experiment allows us to go one step beyond direct questions about preferences for same-gender physicians and to quantify these preferences in terms of an appointment attribute that can have important consequences for health: waiting longer to see a physician. The real-world administrative data allow us to confirm that these preferences are reflected in actual patient behavior and to measure the implications for timing of care.

In our hypothetical choice experiment, run on an online survey platform, we randomized respondents into one of three medical specialties: urology, psychiatry, or orthopedics. These specialties were chosen to reflect variation in the salience and potential importance of gender and other social factors, a feature we call gender sensitivity. Urology is a gender-sensitive specialty in that it relates to an intimate physical area, and can involve physical examination

(something that is made clear to respondents in the survey). Psychiatry could include gender-sensitive discussions, depending on patient history, and involves discussions that can make a patient feel emotionally vulnerable, but no physical exam. Orthopedics, by comparison, is relatively neutral in relating to non-gender-specific physical areas and involving less sensitive communication than psychiatry. Respondents were then asked to repeatedly choose between two physician appointments with randomized appointment attributes, including physician gender, wait time, travel time, cost, and years of physician experience. Respondents also saw patient star ratings (ratings out of five stars), which proxy for physician quality. Such ratings are now commonplace, and often displayed by healthcare systems that collect reviews from verified patients. For example, Stanford Healthcare sends patients a link to a survey after their appointment, and displays the results and associated star ratings on physician profiles. Search engines such as Google will privilege these ratings and display them in search results, given that they are from verified patients, over public sites such as Yelp or Healthgrades.

To measure how patients trade off gender concordance against other attributes, we employ a mixed logit model estimated separately by respondent gender and specialty, allowing preference heterogeneity within each group by treating physician-appointment attribute coefficients as draws from a population distribution. For each group, we calculate willingness-to-pay for gender concordance as the marginal rate of substitution between the gender concordance indicator and appointment cost attributes, specifically wait time (weeks), co-pay (dollars), and travel time (hours). We focus primarily on wait time as our numeraire, given the potential for heterogeneous understanding of its cost arising from the asymmetry of information in healthcare, and subsequent downstream health implications.

We find that a significant gender concordance premium exists for female physicians in some specialties, but by no means all. The strength of the premium varies according to the gender sensitivity of the specialty. Female respondents exhibit meaningful preferences for gender concordance in the gender-sensitive context of urology, where they are willing to wait nearly two additional weeks to see a female physician. Women are similarly willing to wait longer for gender-concordant psychiatrists. But women randomized into the orthopedics scenario, and men in all scenarios, exhibit little to no preference for gender concordance.

We also show that the strength of gender concordance preferences differs by respondents' reported prior experiences. Women who report having previously been treated negatively or unfairly due to gender in a healthcare setting have a higher willingness-to-pay for concordance than those who do not. However, even within the group with no such experiences, the willingness-to-pay for gender concordance in the gender-sensitive specialty, urology, is statistically significantly different than zero. The contrast is sharpest in psychiatry, where women who report prior negative experiences are willing to wait more than two additional weeks for a female psychiatrist, relative to those without such experiences (who exhibit no significant preference).

Finally, we examine whether the preferences observed in the experiment and heterogeneity in access to gender-concordant physicians translate into real-world care-seeking behavior. We test this by analysing real-world data from two different facets of the healthcare system in gender-sensitive specialties: realised physician-patient matching, and patient ratings of physicians. First, using a 20% sample of Medicare claims data from 2006–2018, we study how gender concordance preferences affect patient-physician matching and wait times in the gender-sensitive specialty of urology. To match our hypothetical choice scenario, we focus on

patients who ultimately receive a bladder cancer diagnosis, for whom delays in care may have particularly high stakes. Consistent with the survey findings, female patients are more likely to see female urologists than their male counterparts. As the supply of female urologists increases in a Hospital Referral Region (HRR, regional markets defined by referral patterns for tertiary medical care), all patients are more likely to see female urologists, but the gradient is far steeper for women, suggesting their ability to act on gender-concordant preferences is constrained by the limited supply of female urologists. We find that female patients who see a female urologist wait over 10% longer than those who see a male urologist, with no corresponding gap for male patients who see a female urologist. These patterns suggest that preferences for gender-concordant physicians among female bladder cancer patients delay access to care. Second, using data on online patient ratings of physicians, we compare patient reviews for female versus male physicians in gender-sensitive specialties: obstetrics and gynecology (OB/GYN), and urology. Given we observe strong preferences for gender concordance among female respondents in gender-sensitive specialties, we expect that gender *dis*cordant physicians in such specialties have to compensate via quality. We observe no such preference among male respondents. So, in OB/GYN, given the population is entirely female, we expect a divergence in patient ratings by physician gender, with male OB/GYNs having higher ratings. In urology, where the patient population is mostly male, there should be no such divergence. We document this comparison in data on over 40,000 physicians.

This paper contributes to several strands of literature. The most directly related one concerns the economics of gender-concordant matching in healthcare markets. Reyes (2006; 2008) shows that excess demand for female OB/GYNs clears through both monetary and non-monetary prices: female physicians in that specialty charge higher fees and face longer wait times, with the relative adjustment depending on the rigidity of price-setting institutions. McDevitt and Roberts (2014) explore imperfect competition in the market for urologists, showing that practices strategically differentiate themselves by employing female urologists. Markets without a female urologist have significantly higher bladder cancer mortality among female patients, a finding we expand on using both behavioral and granular appointment-level evidence. Walker et al. (2024) combine survey and field experiment data to show that female primary care physicians face excess demand, with female patients preferring them to male physicians by over 50 percentage points, resulting in longer wait times among female primary care physicians. Lamiraud et al. (2026) find that in France, in areas with a relative scarcity of gender-concordant primary care physicians of either gender, increasing the supply of gender-discordant physicians does not increase care-seeking to the under-served gender. Our paper contributes to this literature by providing the first direct experimental estimates of willingness to pay for concordance across multiple specialties, describing a key dimension of variation in the preferences (gender sensitivity of specialty), and linking those estimates to real-world care-seeking patterns.

A second strand examines how patient–physician demographic concordance affects health outcomes and the mechanisms behind it. Alsan, Garrick and Graziani (2019) use a randomized field experiment to show that Black men are more likely to accept preventive care when assigned a Black physician, with effects concentrated among patients with the highest initial distrust of medicine. Greenwood, Carnahan and Huang (2018) find higher in-hospital mortality among female heart attack patients treated by male physicians, Wallis et al. (2022) document worse surgical outcomes for female patients with male surgeons, and Tsugawa et al. (2017)

and Miyawaki et al. (2024) find lower mortality among elderly Medicare patients of female internists, with the benefit being largest for female patients. Our paper suggests that if patients actively seek concordant providers, and are willing to wait to do so, then the health benefits of concordance documented in this literature may be unevenly distributed depending on local access to concordant physicians.

A third strand of literature examines how the proliferation of online physician-search platforms has fundamentally altered the market for physician selection (Chen and Lee, 2024; Chan, 2026).[1] Prior to the emergence of platforms such as Healthgrades, Zocdoc, and Yelp, patients relied on a narrow set of information channels, such as physician referrals, word of mouth, and advertising, that offered limited coverage and little ability to compare providers on observable attributes. Online platforms have since expanded both the choice set available to patients and the richness of information at their disposal, enabling comparisons across physician experience, credentials, patient ratings, and narrative reviews at low search cost. Some of these platforms allow users to preemptively filter on physician characteristics, such as gender or language. While in the past, patients may have preferred a physician with certain demographic characteristics, they had limited ability to act on these preferences without provoking social awkwardness. Chan (2026) shows that when patients select physicians through online platforms, they will discriminate against Black and Asian physicians without the presence of a quality signal, such as patient ratings, suggesting that physician demographics and the informational design of these platforms affects how patients choose care. In this paper, we show how these online tools could alter patient-physician matching by enabling patients to act directly on preferences for concordance.

The remainder of the paper proceeds as follows. Section 2 documents key contextual facts relevant to the choice we study, including variation in the gender composition of physicians across specialties and regions, and the growth in use of online tools for physician search. Section 3 describes our hypothetical choice experiment design and empirical methodology. Section 4 describes the results of the experiment. Section 5 links our findings to real-world behavior using Medicare data, and Section 6 tests a further prediction of concordance preferences using online physician ratings. Section 7 concludes.

2. Background

In this section, we document two descriptive facts that relate to the patient search process for physician-appointments, to better contextualize the choice environment we study. The first is substantial geographic heterogeneity in physician gender composition and in the availability of female physicians across specialties; for example, 17% of all HRRs have no female urologists. When combined with gender concordance preferences, this heterogeneity could lead patients to wait longer to see a physician, particularly in areas with limited or no access to gender-concordant providers. We then show that the use of online search tools to find health information, search for physicians, and book appointments has grown substantially over time, meaning the patients are increasingly able to act on the preferences our survey measures.

[1] Similar changes have occurred in the market for other healthcare providers, such as nursing homes (Werner et al., 2012) or hospitals (McCarthy, Sanbower and Sánchez Aragón, 2024).

2.1 Physician Gender by Specialty and Area

To better understand the accessibility of male and female physicians by area and clinical specialty, we make use of the publicly available National Plan and Provider Enumeration System (NPPES) data file for September 2025, obtained from the National Bureau of Economic Research (NBER). The NPPES is maintained by the Centers for Medicare and Medicaid Services (CMS) and contains records for every healthcare provider issued a National Provider Identifier (NPI) in the US, including their gender, practice location, and one or more Healthcare Provider Taxonomy codes that encode their type and specialty. We use the ZIP from practice location to infer HRR, via a ZIP-to-HRR crosswalk published by the Dartmouth Atlas.

Starting from 9,129,558 total NPI records, we restrict the sample to individual providers (dropping organizational entities), remove deactivated NPIs, and then retain only records whose primary taxonomy code maps to a physician specialty, which excludes nurses, therapists, pharmacists, and other non-physician providers.[2] A further 330 records are dropped for missing gender, and 14,907 for practice ZIP codes that do not match the HRR crosswalk (primarily APO/FPO military addresses and US territories). The final analytic file contains 1,189,697 active physicians with known gender, specialty, and HRR location.[3]

There is substantial variation in the share of female physicians across medical specialty, as shown in Figure 1. Relatively few specialties have a majority female workforce; only pediatrics, obstetrics and gynecology, and dermatology, and none are above 70% female. In other specialties, such as urology and orthopedics, less than 20% of physicians are female.

This creates variation in access to female and male physicians across the US. We map the share of physicians who are female across HRRs for four specialties in Figure 2.[4] For urology and orthopedics, the median share of female physicians by HRR is around one in ten; for obstetrics and gynecology and psychiatry, it is four to six in ten. Specialties with a lower share of female physicians result in some HRRs having no female physicians for that specialty. Indeed, this is a relatively common occurrence in urology and orthopedics, two of the specialties that enter hypothetical choice scenarios in our survey. Out of 306 HRRs, 53 (17.3%) have no female urologist, and 30 (9.8%) have no female orthopedist.[5]

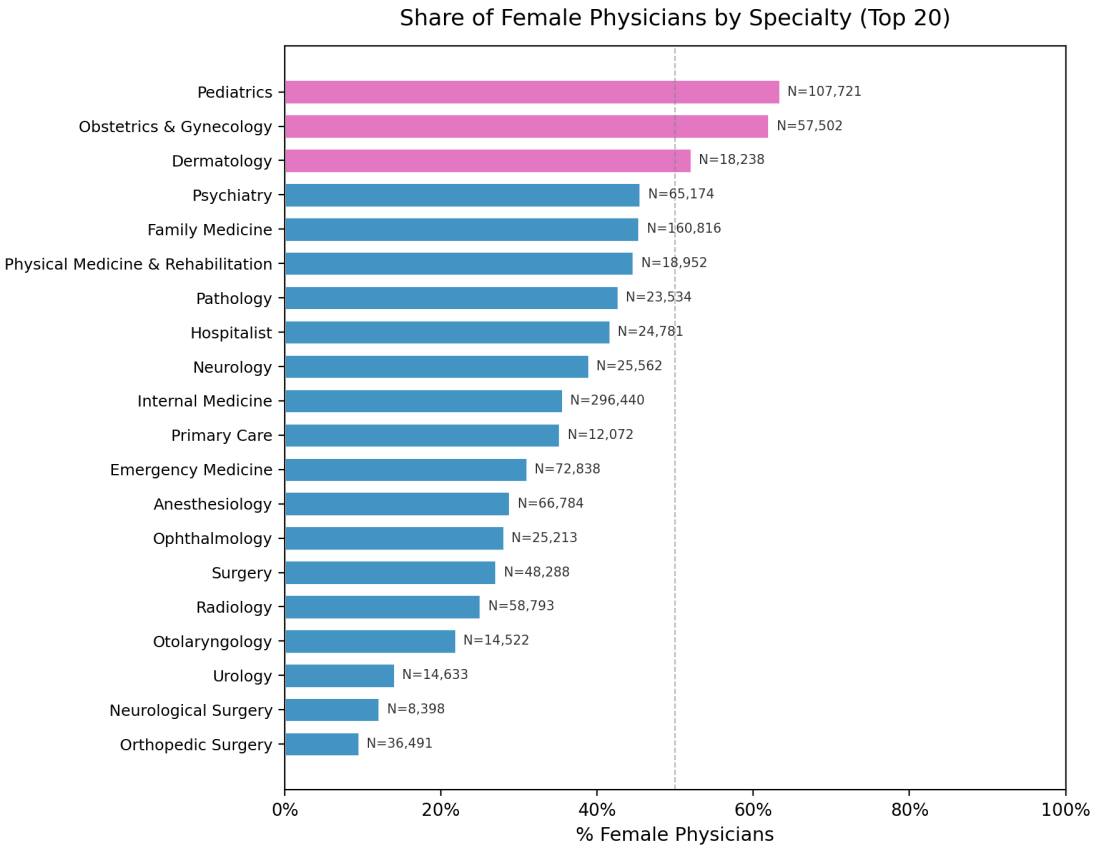
[2]Each NPI record in the NPPES file may list up to 15 Healthcare Provider Taxonomy codes, reflecting the range of services a provider is qualified to deliver. One of these codes can be designated as the provider's "primary" taxonomy via a flag field (pprimtax). We assign each physician a single specialty using the first taxonomy code flagged as primary; if no primary flag is set, we default to the first-listed taxonomy code. Among the 1,204,934 active physicians in our sample, 37.5% list two or more taxonomy codes. Of these physicians, 68.6% have their primary flag on the first taxonomy slot, 21.9% on the second, and 6.3% on the third. The remaining 3.2% of the sample (38,982 physicians) have no primary flag set on any of their first three codes and are assigned the specialty corresponding to their first-listed taxonomy code. We map each taxonomy code to a specialty category (e.g., "Internal Medicine", "Psychiatry", "Orthopedic Surgery") using the NUCC Healthcare Provider Taxonomy Code Set classification level, which groups subspecialties under their parent discipline. Each physician is counted exactly once under a single specialty.

[3]A summary of these sample restrictions can be seen in Table A1.

[4]We show which HRRs lack access to any female physician (by the same specialties) in Figure A1.

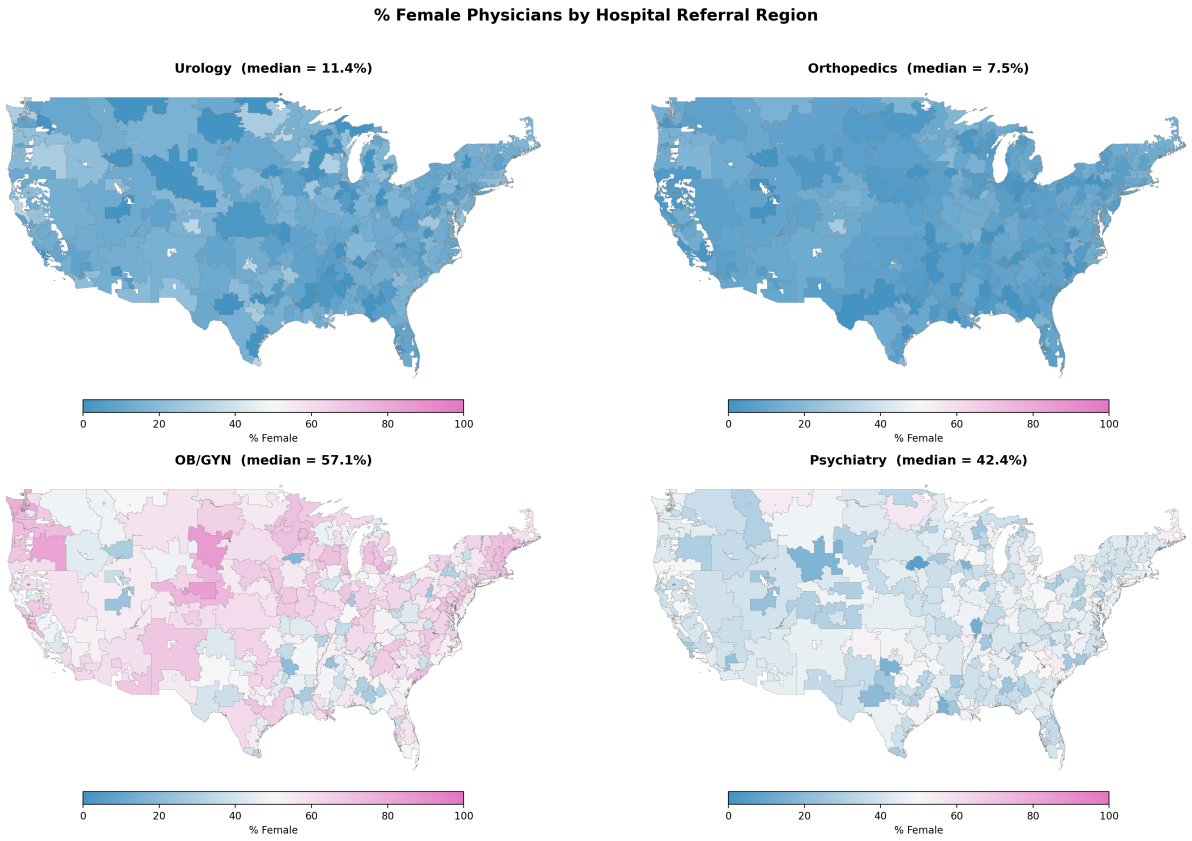
[5]We show the significant geographic dispersion of these HRRs across the US in Figure A1.

Figure 1: Gender by Medical Specialty



Notes: Data from NPPES NPI file September 2025, downloaded from NBER. Analysis restricted to top twenty specialties by number of physicians. (Internal medicine includes gastroenterology, cardiology, nephrology, pulmonology, rheumatology, and other subspecialties.) Specialty inferred from provider taxonomy codes.

Figure 2: Share of Female Physicians by Hospital Referral Region and Specialty



Notes: Data from NPPES NPI file September 2025, downloaded from NBER. Specialty inferred from provider taxonomy codes. Practice location linked to HRR via ZIP, using Dartmous Atlas crosswalk.

2.2 Online Search for Health Information

We next provide background on how people are using online tools to search for health information, including physician profiles. To do so, we use the Health Information National Trends Survey (HINTS), a nationally representative, repeated cross-sectional survey of US adults run by the National Cancer Institute (National Cancer Institute, 2025). The purpose of HINTS is to measure "the American public's knowledge of, attitudes toward, and use of cancer- and health-related information," and to "monitor changes in the rapidly evolving fields of health communication and health information technology and to create more effective health communication strategies across different populations."

We use survey cycles from 2008 through 2024. The exact wording of questions changed over some cycles, and we group them where appropriate. We focus on three relevant survey topics: (1) Did the respondent *look for a provider* online? (fielded 2008-2017); (2) Did the respondent *make an appointment with a healthcare provider* online, or using a computer, smartphone, or other electronic means? (fielded in 2017-2024); (3) Did the respondent use the Internet or electronic means to *look for health information*? (fielded 2011-2024).

The share of adults reporting any of these three behaviors has risen substantially in the past two decades, as shown in Figure 3a. Looking for health information and making appointments online have both become majority behaviors. Particularly relevant for our context: by 2024, two-thirds of US adults reported making healthcare appointments online. And, by 2017 (after which the survey item was discontinued), over a third of US adults said they had searched for a healthcare provider online, up from a quarter not ten years previously.

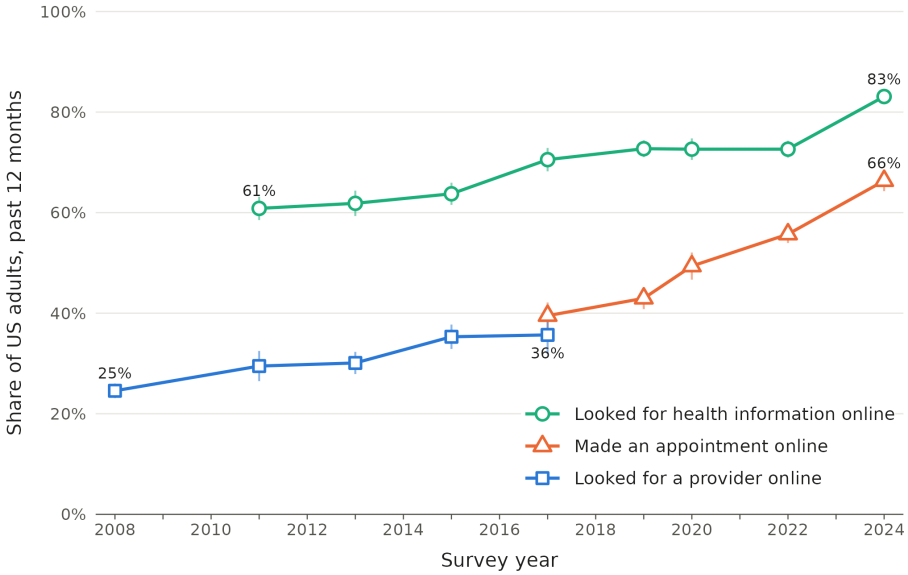
Split by gender, rates of usage of online tools for searching for a healthcare provider or making an appointment are broadly similar. Figure 3b-3c show that in most years women have slightly higher rates of use. We also report the split by age, in Appendix Figure A2. Younger age groups report higher rates of usage for healthcare purposes, but even among adults age 65 and over, nearly half had used online tools to make healthcare appointments in 2024. Overall, as online search and booking have become widespread, patients have gained more opportunities to observe physician characteristics before an appointment and to choose providers who match their preferences, including preferences for gender concordance.

3. Experimental Design & Methodology

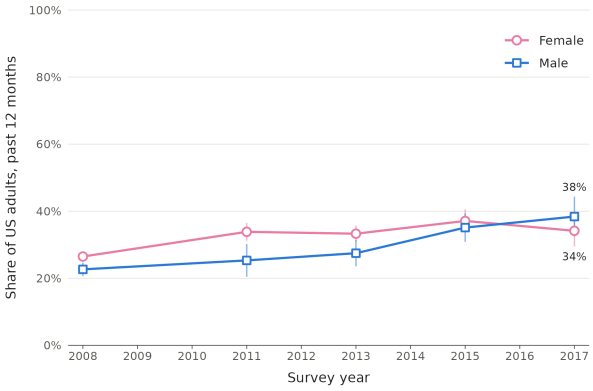
In this section, we outline our experimental design and associated methodology. Our experimental design is, in essence, a hypothetical choice (or conjoint) experiment, a method of growing use in economics (Wiswall and Zafar, 2018; Gallen and Wasserman, 2023; Andrew and Adams, 2025; Chan, 2026). The design uses randomized variation in physician attributes across a large sample of participant choices between physician appointments in three medical specialties, enabling us to estimate willingness-to-pay for gender concordance using a discrete choice model. Though this tool relies on responses collected outside of a real healthcare context, it provides numerous advantages over observational data. First, by designing the choice environment within the survey, we have control over the respondent's information and choice sets – two key objects usually unobservable in observational healthcare data. Second, this design control means that respondents can be presented with gender-concordant physicians

Figure 3: Use of online tools for health care

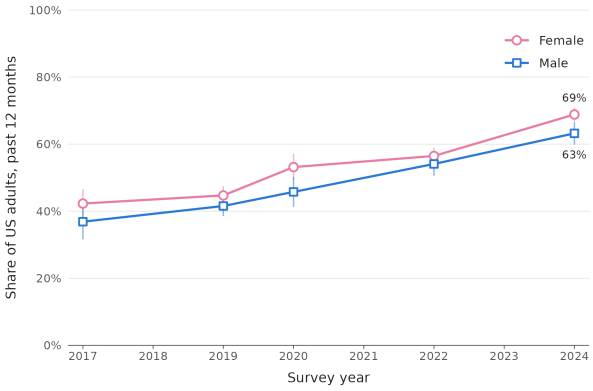
(a) Share of US adults reporting use of online tools for health purposes



(b) Looked for a health care provider online



(c) Made an appointment with a provider online



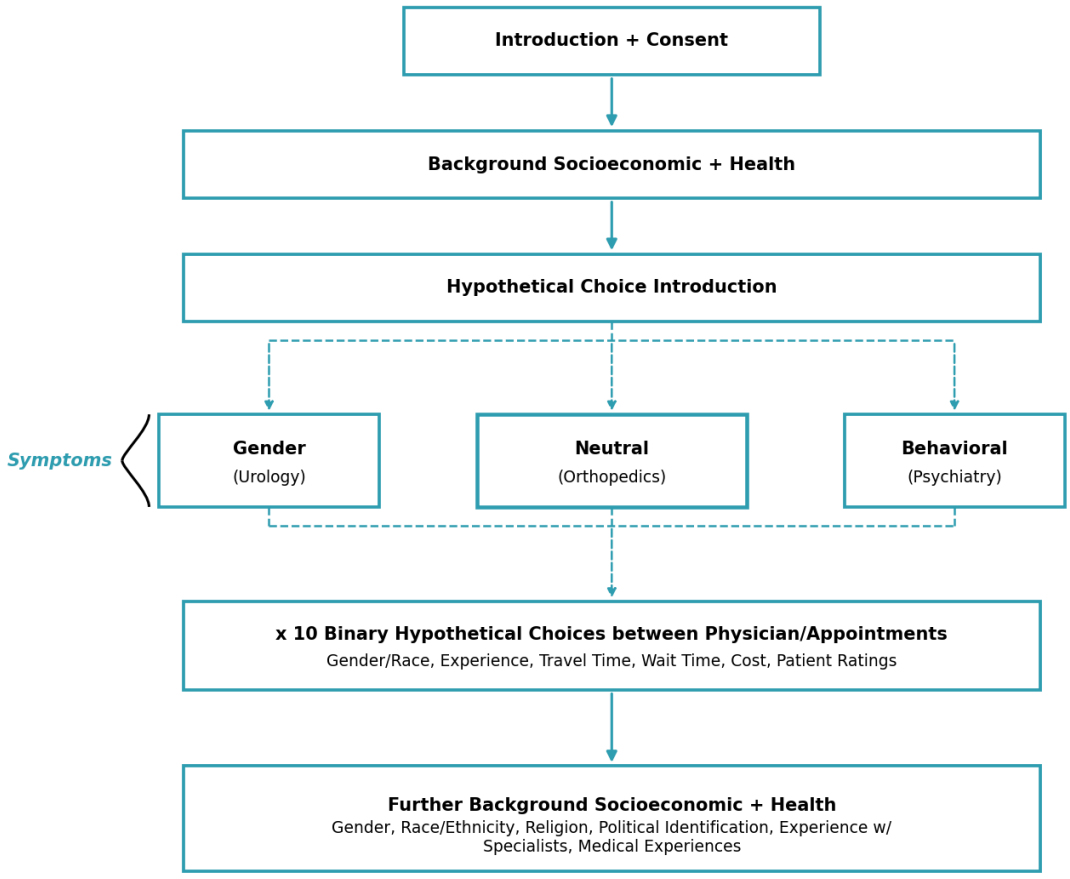
Notes: Figure (a) reports survey-weighted shares of all US adults. Figures (b) and (c) report survey-weighted shares of women and men who did each activity in the past 12 months; the denominator is all adults of that gender, including those who do not use the internet. Vertical bars are 95% confidence intervals from 50 jackknife replicate weights. In panel (a), the provider item (2008–2017) and the information item in 2011–2015 ask about "the Internet"; the 2017–2020 items ask about "a computer, smartphone, or other electronic means"; and the 2022–2024 items return to "the Internet." The 2017 provider point is from the HINTS-FDA cycle, which keeps the "Internet" wording. The provider item was not asked after 2017. In panel (b), the 2017 point is from the HINTS-FDA cycle, which keeps the 2008–2015 "Internet" wording. In panel (c), the 2017–2020 item asks about "a computer, smartphone, or other electronic means"; the 2022–2024 item asks about "the Internet." Source: HINTS public-use files.

as options, something that does not exist in many healthcare markets, as documented above. Finally, the tool affords us the opportunity to examine the drivers of preferences (such as prior experience in healthcare) in more detail than would be possible in an active healthcare environment. Our approach is close to Chan (2026), who presents customers of an online physician-booking platform with menus of hypothetical doctors described by price, distance, a blurred profile photograph that conveys race and gender, and, in some cases, patient ratings, in a layout mimicking the platform's real search results. He validates the choice-based conjoint format in this setting, showing that preferences estimated from the hypothetical menus are strong predictors of respondents' subsequent bookings of physician-appointments.

3.1 Experimental Design

Our survey allows us to estimate willingness-to-pay for gender concordance, relative to other physician-appointment attributes. A diagram of the survey flow is available in Figure 4.

Figure 4: Hypothetical Choice Experiment Survey Flow



Respondents entered the survey, consented, and passed an instructional manipulation check to assess attentiveness. We then asked a range of background socioeconomic questions

(e.g., age, household income) and health questions (e.g., chronic conditions, prescriptions) to gauge their experience interacting with the healthcare system.

We then explained to respondents that we were going to present them with a hypothetical scenario in which they have developed some symptoms and are seeking medical advice. In each scenario, a primary care physician refers them to a specialist in a more or less-gender-sensitive specialty. We told respondents that we would then ask them to choose between potential physicians and appointments.

We chose three specialties and three appointment types, in consultation with physicians and the medical literature, to achieve variation in potential sensitivity to concordance, while maintaining an identical experimental environment across respondents. For example, if we had chosen a gynecological appointment for female respondents and a urological appointment for male respondents randomized into the more gender sensitive specialty, or screening for breast versus prostate cancer, our comparison of the two groups on preferences for gender concordance would have been clouded by an experimental treatment that differed on many dimensions. We chose urology, psychiatry, and orthopedics for the variation they offer in sensitivity to concordance, and their familiarity to respondents relative to other specialties that treat lower incidence conditions.[6] Urology is gender-sensitive in both a social and physical sense, given it involves a gendered area of the body and, usually, intimate physical exam in an appointment. Psychiatry is gender-sensitive in a behavioral sense, in that it involves intimate discussion that can tread on emotional and gender-sensitive topics. Orthopedics is more neutral, in that it often pertains to non-gendered areas of the body (e.g., hip, knee), for which a physical exam is unlikely to be gender-sensitive, and in which diagnostic techniques are less intimate (e.g., imaging modalities such as MRI or X-ray).

Within each specialty, we chose symptoms for conditions that would lead to similar initial appointments for all patient types, that again varied in their sensitivity to concordance, and which could be easily explained. Symptom descriptions were drawn from a range of sources, such as the CDC, Mayo Clinic, and United Kingdom's NHS. We outlined three to four symptoms to respondents, and provided a time horizon over which the symptoms have been occurring. Then, we explained that in the scenario the patient's primary care physician has referred them to see a specialist. We explained what the specialist's role is and what would take place at the appointment. This was done primarily for realism; a primary care physician would likely explain that he/she was referring the patient to a specialist, what the specialty is, and what diagnostic tests the specialist might do at the appointment. We provide more detail on the information the respondents are offered in Appendix B.

After the information treatments, we gave respondents a preview of the hypothetical choice. We told respondents the exact number of choices we would ask them to make and emphasized that they should choose as they would for themselves. We then outlined the characteristics of the physician-appointments that they would be shown, and explained each one. Critically, we told respondents to assume that the physicians and appointments were otherwise identical in other aspects; implicitly, choices should be made based *only* on variation in the characteristics we outline. Respondents then made ten choices across

[6]Other specialties that may appear neutral in that they do not pertain to a gendered body part, such as dermatology, were discounted because initial appointments would likely involve an intimate examination such as a skin mapping exercise.

physician-appointments. Figure 5 provides an example of the type of choice they were offered within the survey environment.

We presented respondents with ten choices, each between two physician-appointments. The respondents were able to choose neither appointment, which serves as our outside option.

Our experiment has three key aspects of variation to help us estimate a range of relevant willingness-to-pay metrics.

1. **Variation in medical setting** – Respondents are randomly assigned to one of three medical areas with associated symptoms for an underlying condition (urology, orthopedics, or psychiatry), as outlined above.

2. **Variation in physician gender** – We vary physician gender randomly for each physician profile (20 total, over 10 choices). We decided the probability of each gender with two goals: first, approximate the share of female physicians in the true distribution and so have a plausible share of both genders; second, generate a sufficient number of choices where both options were available. We chose a 60% chance of a male physician and a 40% chance of a female physician. This is close to the true distribution over all specialties, while also allowing for a higher number of observed choices between physicians of different genders.

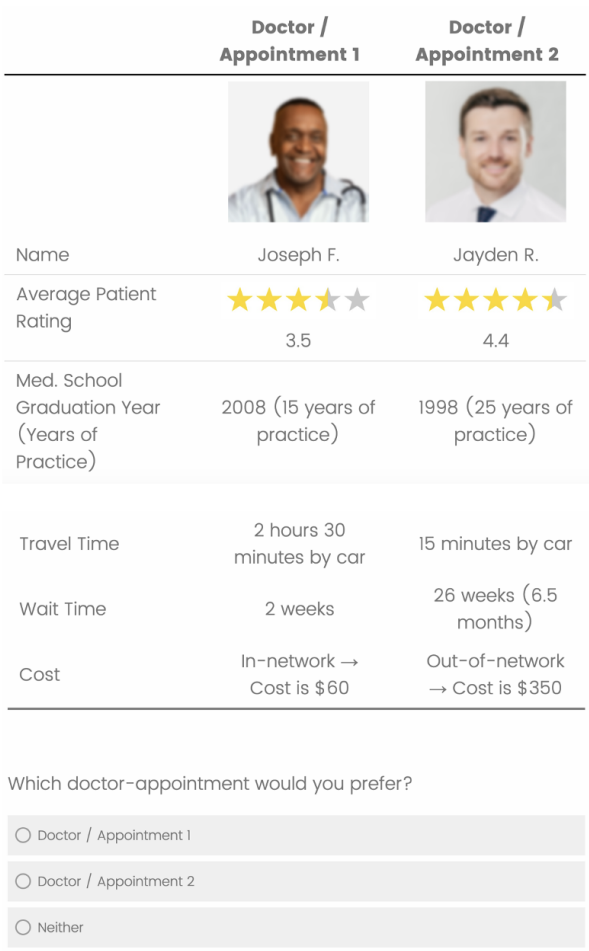
3. **Variation in numéraires** – Patients can pay for concordance via various numéraires: money, travel time, experience, patient ratings, or wait time. We included all of these in each physician profile, but we will focus on wait time as our primary numéraire.

To design the hypothetical choice, we reviewed online search tools for physicians—both independent and those built internally to several major healthcare systems (academic and non-academic)—to mimic the real-world process of choosing a physician as closely as possible. The information offered in these search tools heavily influenced the information we chose to include (and exclude) in our hypothetical choice. Generally, we found that most healthcare systems presented a physician's photo and name, as well as their specialty, clinic/location, and often a patient star rating. Age could often be roughly inferred from photos, and often more detail on experience and education was available by clicking on physician profiles. Some search tools presented appointment availability and distance from the user's location (if the user entered a ZIP code). Others allowed users to filter results by physician gender, giving preferences for gender concordance an opportunity to impact not only the choice of physician, but the choice set overall. We provide screenshots of some of the online search tools we looked at in Appendix B.3.

Influenced by these examples, we chose photos and names as our preferred method of conveying gender. This has the benefit of providing greater realism to physician search online, as well as avoiding highlighting gender in the physician choice, potentially creating experimental demand effects. Following the experimental literature, we chose to blur these photos (Fong and Luttmer, 2009; Chan, 2026). Doing so largely preserves the ability to infer gender, but creates enough noise in the image that other hard-to-observe factors influencing physician choice (such as "attractiveness", professionalism, or just the quality of the photo) are made somewhat uniform across images. We used stock images that returned for the search term "physician" or "doctor". Many of these images are from photoshoots designed to

generate stock images of people as physicians, and we select images such that there is greater uniformity in photo quality and composition.

Figure 5: Example of Hypothetical Choice Between Physician-Appointments



The survey finished by eliciting further background and health information. Critically, we only ask for respondent gender *after* the hypothetical choices have been made, to avoid making this characteristic more salient to respondents. We also asked explicitly if respondents had ever had a negative medical experience, and if they perceived that as being due to their gender. Respondents were also offered the chance to give any feedback on the survey, and we asked for their guess as to what the purpose of the survey was.

We fielded our survey on the online platform Prolific between March and April 2026. We aimed for 2,800 respondents. Prolific identified respondents who passed the initial attention check of the survey, were not flagged as potentially using LLMs or bots by the survey platform, and did not complete the survey exceptionally fast. Further limiting the sample to those who completed the survey and did not experience technical errors yielded 2,793 respondents, and 27,930 choice tasks. Demographic characteristics by treatment groups are presented in Table 1. The three groups are balanced across different demographic characteristics, such as age, gender, race, education, marital status, employment, and health status, among others.

Table 1: Demographic characteristics by treatment group (completes, $n = 2793$)

	Orthopedics ($n = 935$)	Urology ($n = 943$)	Psychiatry ($n = 915$)	Test	
Age, mean (SD)	42.6 (13.2)	42.9 (13.4)	42.6 (14.0)	$F = 0.14$	
Gender					
Male	463 (49.5%)	473 (50.2%)	470 (51.4%)	$\chi^2 = 0.65$	
Female	472 (50.5%)	470 (49.8%)	445 (48.6%)		
Race					
White	468 (50.1%)	469 (49.7%)	477 (52.1%)	$\chi^2 = 1.25$	
Black	467 (49.9%)	474 (50.3%)	438 (47.9%)		
Education					
Less than HS	6 (0.6%)	7 (0.7%)	9 (1.0%)	$\chi^2 = 2.45$	
High school	119 (12.7%)	126 (13.4%)	111 (12.1%)		
Some college	192 (20.5%)	185 (19.6%)	171 (18.7%)		
College+	618 (66.1%)	625 (66.3%)	624 (68.2%)		
Marital status					
Single	446 (47.7%)	458 (48.6%)	421 (46.0%)	$\chi^2 = 1.00$	
Married	385 (41.2%)	380 (40.3%)	384 (42.0%)		
Separated/divorced	88 (9.4%)	85 (9.0%)	84 (9.2%)		
Employment					
Employed	726 (77.6%)	731 (77.5%)	728 (79.6%)	$\chi^2 = 1.42$	
Unemployed	209 (22.4%)	212 (22.5%)	187 (20.4%)		
Insurance					
Insured	926 (99.0%)	933 (98.9%)	901 (98.5%)	$\chi^2 = 1.04$	
Uninsured	8 (0.9%)	9 (1.0%)	12 (1.3%)		
Health status					
Chronic condition	437 (46.7%)	439 (46.6%)	434 (47.4%)	$\chi^2 = 0.16$	
Prescription medications	434 (46.4%)	421 (44.6%)	436 (47.7%)	$\chi^2 = 1.71$	
Emergency room use	750 (80.2%)	782 (82.9%)	718 (78.5%)	$\chi^2 = 6.00$	*
Religion					
Christian	503 (53.8%)	538 (57.1%)	511 (55.8%)	$\chi^2 = 6.35$	
Jewish	10 (1.1%)	9 (1.0%)	13 (1.4%)		
Muslim	9 (1.0%)	15 (1.6%)	16 (1.7%)		
Atheist/Agnostic/None	384 (41.1%)	350 (37.1%)	346 (37.8%)		
Other	29 (3.1%)	31 (3.3%)	28 (3.1%)		
Political affiliation					
Republican	198 (21.2%)	195 (20.7%)	211 (23.1%)	$\chi^2 = 8.40$	
Democrat	432 (46.2%)	435 (46.1%)	417 (45.6%)		
Independent	254 (27.2%)	265 (28.1%)	259 (28.3%)		
Other/no preference	51 (5.5%)	48 (5.1%)	28 (3.1%)		

χ^2 test for categorical; one-way ANOVA F-test for continuous.

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

3.2 Methodology

Our estimation strategy is a mixed logit model estimated separately by respondent gender and physician specialty. The mixed logit setup allows for preference heterogeneity within each group g by treating a subset of coefficients as random coefficients drawn from a population distribution. Let i index respondents, j index physician-appointment options within choice task t, and $g \in \{\text{men}, \text{women}\}$ denote the respondent gender group. The utility that respondent i in group g derives from choosing appointment option j in task t is

$$U_{ijt}^{(g)} = X_{ijt}'\beta_i^{(g)} + \epsilon_{ijt}, \tag{1}$$

where X_{ijt} is a vector of observed physician and appointment characteristics, $\beta_i^{(g)}$ is respondent i's vector of marginal utilities over these characteristics, and ϵ_{ijt} is an i.i.d. Type I Extreme Value choice-specific utility shock. The vector X_{ijt} includes co-pay (in dollars), wait time (in weeks), travel time (in hours), physician experience (in years), patient ratings (in stars), an indicator for the outside option, and $\text{GenderConcordant}_{ijt}$, which equals one when the physician and respondent share the same gender.

To allow for preference heterogeneity within each group, we assume:

$$\beta_i^{(g)} = \mu^{(g)} + \eta_i^{(g)}, \qquad \eta_i^{(g)} \sim \mathcal{N}\left(0,\, \Sigma^{(g)}\right), \tag{2}$$

where $\mu^{(g)}$ is the vector of population mean utilities for group g and $\Sigma^{(g)}$ is a diagonal covariance matrix governing within-group preference dispersion. Preference heterogeneity is particularly relevant for the gender concordance, wait time, co-pay, travel time, and patient ratings. For example, we document below that preferences for gender concordance vary with respondents' prior healthcare experiences. As such, we allow the coefficients on gender concordance, wait time, co-pay, travel time, and patient ratings to vary across respondents, while holding the coefficients on physician experience and the outside option fixed. This is done for computational parsimony, and allowing coefficients on experience to vary across respondents does not affect results.

Conditional on respondent i's coefficient vector $\beta_i^{(g)}$, the probability of choosing option j in task t is

$$p_{ijt}(\beta_i^{(g)}) = \frac{\exp\left(X_{ijt}'\beta_i^{(g)}\right)}{\sum_{k \in C_{it}} \exp\left(X_{ikt}'\beta_i^{(g)}\right)}. \tag{3}$$

Because each respondent completes ten choice tasks, we hold $\beta_i^{(g)}$ fixed across their tasks. The probability of respondent i's observed sequence of choices is therefore obtained by integrating the product of the conditional choice probabilities over the distribution of random coefficients:

$$P_i^{(g)} = \int \left[\prod_{t=1}^{10} \prod_{j \in C_{it}} p_{ijt}(\beta)^{\mathbf{1}\{Y_{it}=j\}} \right] f\left(\beta \mid \mu^{(g)}, \Sigma^{(g)}\right) d\beta. \tag{4}$$

We approximate this integral using 500 Halton draws and estimate $\mu^{(g)}$ and $\Sigma^{(g)}$ by maximizing the simulated log likelihood

$$\mathcal{L}^{(g)} = \sum_{i \in g} \log \widehat{P}_i^{(g)}, \tag{5}$$

where $\widehat{P}_i^{(g)}$ is the simulated approximation to the sequence probability in Equation (4). The estimated mean coefficients $\mu^{(g)}$ capture average marginal utilities within group g, while the estimated standard deviations $\sqrt{\Sigma_{kk}^{(g)}}$ capture within-group heterogeneity in preferences for each random attribute k.

To express preferences for gender concordance in interpretable units, we calculate willingness-to-pay as the marginal rate of substitution between gender concordance and three measures of access costs: wait time, co-pay, and travel time:

$$\text{WTP}^{(g)}_{\text{Gender, Wait}} = -\frac{\mu^{(g)}_{\text{GenderConcordant}}}{\mu^{(g)}_{\text{WaitTime}}} \tag{6}$$

$$\text{WTP}^{(g)}_{\text{Gender, Cost}} = -\frac{\mu^{(g)}_{\text{GenderConcordant}}}{\mu^{(g)}_{\text{CoPay}}} \tag{7}$$

$$\text{WTP}^{(g)}_{\text{Gender, Travel}} = -\frac{\mu^{(g)}_{\text{GenderConcordant}}}{\mu^{(g)}_{\text{TravelTime}}}. \tag{8}$$

These measures are evaluated at the mean utility coefficients for group g, rather than averaged over respondents' individual willingness-to-pay ratios. They can be interpreted as the additional weeks of waiting, dollars of co-pay, or hours of travel associated with gender concordance relative to a non-concordant physician, holding other attributes constant. We obtain standard errors using the delta method. Throughout, we focus on willingness-to-pay expressed in weeks of waiting, as wait time is a salient and clinically meaningful measure of access costs. Corresponding estimates in dollars of co-pay and hours of travel are also reported.

4. Experimental Results

We present the experimental results in three parts. We first examine stated preferences for same-gender physicians, which confirm that women report stronger concordance preferences than men. We then estimate willingness-to-pay for gender concordance from the hypothetical choice experiment, finding that women are willing to wait longer to see a female physician in more gender-sensitive specialties, while men exhibit little to no concordance preference across specialties. We show that these preferences are stronger among women who report prior negative healthcare experiences due to gender.

4.1 Stated Preferences

Before turning to the results of the hypothetical choice experiment, we first examine whether respondents report preferences for same-gender physicians, and whether these differ by

respondent gender. These questions were asked after the choice tasks had concluded, so they do not contaminate the experiment itself.

We find that 34.3% of respondents report a preference for same-gender physicians, independent of specialty. compared to only 6.4% who report a preference for different-gender physicians. There are pronounced gender differences, which we outline in Table 2. While 51.3% of women report preferring a same-gender physician, only 17.7% of men do, a difference of 33.6 percentage points that is statistically significant. At the same time, only 2.8% of women report preferring different-gender physicians, compared to 10% of men, suggesting that female physicians may be preferred overall.

Table 2: Stated preferences and beliefs by gender ($N = 2793$)

	Male ($n = 1406$)		Female ($n = 1387$)		Diff.	
	n	%	n	%	(pp)	
Gender concordance preference						
Prefer own gender	249	17.7	711	51.3	+33.6	***
Prefer other gender	140	10.0	39	2.8	-7.1	***
No preference	1017	72.3	637	45.9	-26.4	***
Negative experience (gender)						
Yes	103	7.3	428	30.9	+23.5	***
No	1303	92.7	959	69.1	-23.5	***
Trust in medical professionals						
Most can be trusted	1289	91.7	1215	87.6	-4.1	***
Most cannot be trusted	117	8.3	170	12.3	+ 3.9	***

Two-proportion z-test: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

These questions also reveal substantial differences in trust in medicine and reported gender-related experiences. Women are substantially more likely to report negative experiences in healthcare due to gender (30.9% versus 7.3%, a 23.5 percentage point gap) and report lower trust in medical professionals than men (87.6% versus 91.7%, a 4.1 percentage point gap).

The overall takeaway from this end-of-survey questionnaire is that, if patient behavior maps to stated preferences, we would expect, on average, weak to no concordance preferences among men, and substantial preferences for concordance among women.

4.2 Hypothetical Choice Experiment

Willingness-to-Pay across Denominators, Specialties, and Gender We next analyze the results from the hypothetical choice experiment. Table 3 presents willingness-to-pay estimates across different denominators (wait time, cost, travel time) and for different appointment attributes. We find, as expected, that respondents value lower wait time, lower cost, lower travel time, more experience, and higher patient ratings.

A striking pattern emerges when examining how the three access-cost attributes—wait time, cost, and travel time—are valued relative to one another. The dollar value of travel

time is essentially constant across specialties and genders, hovering around $71–$91 per hour (Panel B). This consistency is intuitive, as the opportunity cost of time spent traveling reflects stable individual time preferences that do not vary with the clinical reason for the visit.

Wait time, by contrast, is priced very differently depending on specialty and respondent gender. In dollar terms, men in urology are willing to pay $29 to avoid a week's wait, compared to $12 in psychiatry, and women are willing to pay $20 in urology compared to $9 in psychiatry. The same divergence appears when wait time is expressed in travel-time terms (Panel C), ranging from 0.35 hours per week for male urology patients to 0.15 hours for male psychiatry patients, and 0.25 hours for female urology patients to 0.1 hours for female psychiatry patients. This heterogeneity is consistent with wait time carrying specialty- and symptom-specific costs beyond the pure opportunity cost of time. Perceived urgency and consequences of diagnostic delay plausibly vary across clinical settings in ways that the cost of a car ride does not.

An extra star in patient ratings generates substantial willingness-to-pay across groups, but its valuation varies in interesting ways across numéraires and specialties. Men in the urology arm are willing to pay $209 for an extra star in patient ratings, and would wait 7.3 additional weeks for it. In psychiatry, however, men would wait 12.2 weeks for the same improvement in ratings but pay only $146. This pattern suggests that the implicit exchange rate between dollars and wait time itself varies across specialties, consistent with wait time carrying greater urgency costs in different specialties and presented conditions.

Finally, the table reveals that female respondents are 1.6–2.4 times more likely than male respondents to select the outside option. However, neither women nor men systematically change their likelihood of selecting the outside option as the number in the choice set varies across $\{0, 1, 2\}$.[7] This suggests that respondents did not differentially opt out of choice sets with particular gender compositions.

[7] As shown in Appendix Figure C10.

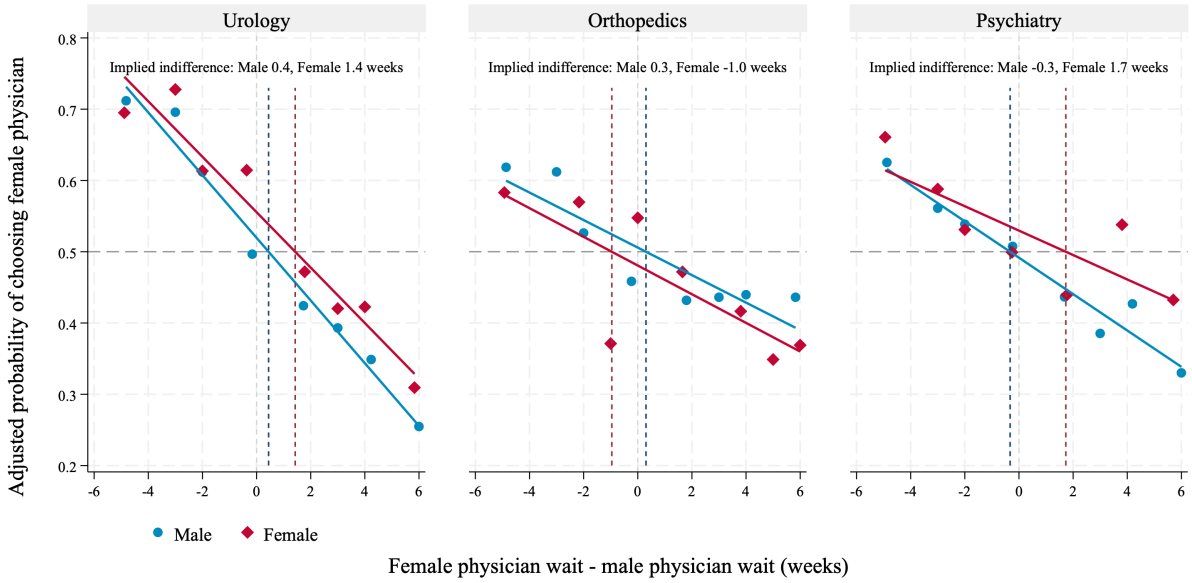
Table 3: Willingness-to-Pay for Different Attributes by Specialty, Gender, and Numéraire

	Urology		Orthopedics		Psychiatry	
	Male	Female	Male	Female	Male	Female
Panel A: WTP in Wait Time (Weeks)						
Cost	-0.035	-0.049	-0.076	-0.081	-0.084	-0.115
	(0.003)	(0.003)	(0.005)	(0.005)	(0.006)	(0.009)
Travel Time	-2.879	-3.976	-5.387	-7.357	-6.748	-9.694
	(0.238)	(0.290)	(0.429)	(0.532)	(0.537)	(0.818)
Experience	0.060	0.026	0.106	0.112	0.174	0.166
	(0.033)	(0.034)	(0.047)	(0.048)	(0.050)	(0.065)
Patient Ratings	7.304	6.080	12.687	12.101	12.225	14.308
	(0.501)	(0.479)	(0.819)	(0.770)	(0.843)	(1.167)
Gender Concordant	-0.074	1.588	-0.309	0.726	-0.841	2.175
	(0.399)	(0.458)	(0.579)	(0.644)	(0.616)	(0.859)
Panel B: WTP in Cost (Dollars)						
Wait Time	-28.648	-20.329	-13.223	-12.348	-11.952	-8.687
	(2.314)	(1.326)	(0.914)	(0.786)	(0.836)	(0.672)
Travel Time	-82.464	-80.831	-71.224	-90.842	-80.655	-84.207
	(8.081)	(6.067)	(5.631)	(6.308)	(6.354)	(6.078)
Experience	1.729	0.523	1.396	1.384	2.077	1.441
	(0.939)	(0.682)	(0.624)	(0.594)	(0.595)	(0.563)
Patient Ratings	209.242	123.595	167.756	149.422	146.115	124.294
	(18.314)	(10.098)	(10.761)	(9.062)	(9.855)	(8.368)
Gender Concordant	-2.109	32.281	-4.086	8.961	-10.056	18.890
	(11.419)	(9.366)	(7.651)	(7.953)	(7.363)	(7.425)
Panel C: WTP in Travel Time (Hours)						
Cost	-0.012	-0.012	-0.014	-0.011	-0.012	-0.012
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Wait Time	-0.347	-0.251	-0.186	-0.136	-0.148	-0.103
	(0.029)	(0.018)	(0.015)	(0.010)	(0.012)	(0.009)
Experience	0.021	0.006	0.020	0.015	0.026	0.017
	(0.011)	(0.008)	(0.009)	(0.007)	(0.007)	(0.007)
Patient Ratings	2.537	1.529	2.355	1.645	1.812	1.476
	(0.230)	(0.132)	(0.181)	(0.115)	(0.143)	(0.110)
Gender Concordant	-0.026	0.399	-0.057	0.099	-0.125	0.224
	(0.139)	(0.117)	(0.107)	(0.088)	(0.092)	(0.088)
Panel D: SD of Random Coefficients						
SD(Gender Concordant)	0.116	0.621	0.077	0.565	0.136	0.492
SD(Wait Time)	0.118	0.115	0.080	0.074	0.086	0.076
SD(Cost)	0.006	0.005	0.006	0.006	0.007	0.007
SD(Travel Time)	0.385	0.422	0.491	0.674	0.764	0.685
SD(Patient Ratings)	0.926	0.853	0.639	0.618	0.727	0.708
Outside option chosen (%)	10.6	23.8	10.6	18.3	13.4	21.7
N of choice tasks	4730	4700	4630	4720	4700	4450
N of respondents	473	470	463	472	470	445

Notes: Willingness-to-pay (WTP) estimates are derived from a mixed logit model estimated separately by respondent gender and medical specialty (urology, orthopedics, and psychiatry). The coefficients on gender concordance, wait time, cost, travel time, and patient ratings are treated as random, drawn from a normal distribution. The coefficients on physician experience and the outside option are fixed. The model is estimated by maximum simulated likelihood using 500 Halton draws. Respondents made 10 discrete choices each between two physician profiles varying in cost, wait time, travel time, physician experience, patient ratings, and gender concordance. Panel A reports willingness-to-pay expressed in weeks of wait time, Panel B in dollars of out-of-pocket cost, and Panel C in hours of travel time. Panel D reports the estimated standard deviations (SD) of the random coefficients, which capture within-group preference heterogeneity. In each panel, WTP for a given attribute equals the negative ratio of that attribute's mean coefficient to the mean numeraire coefficient. Standard errors are obtained via the delta method.

Descriptive Choice Patterns To characterize the choice patterns underlying the gender-concordance estimates before imposing a parametric relationship between relative wait time and physician choice, we first examine how the probability of choosing the female physician varies with her wait time relative to the male physician. We restrict attention to choice tasks in which respondents were presented with exactly one male and one female physician and selected one of the two. Within each specialty and respondent gender, we residualize the indicator for choosing the female physician with respect to differences in cost, travel time, physician experience, and patient ratings, and add back the subgroup mean. Figure 6 plots these adjusted choice probabilities against the female-minus-male difference in wait time.

Figure 6: Choice of Female Physicians by Relative Wait Time



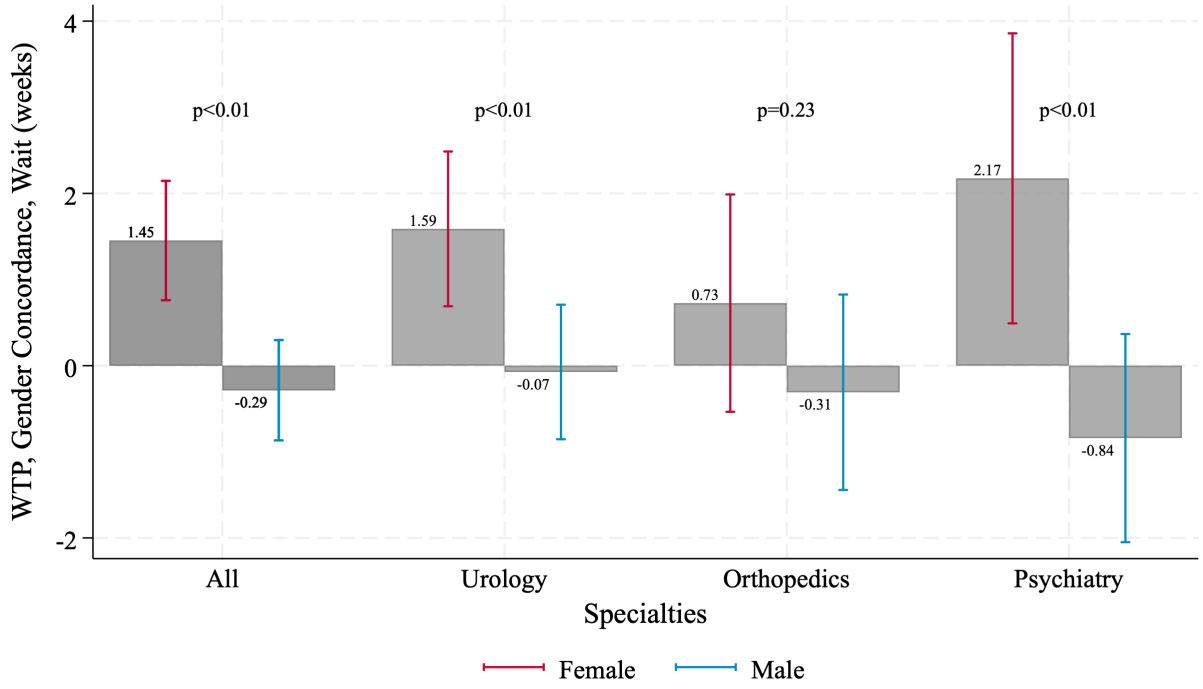
Notes: $N = 2{,}680$ unique respondents and 11,205 choice tasks in which respondents were presented with one male and one female physician and selected one of the two. The figure plots the adjusted probability of choosing the female physician against the difference in wait time between the female and male physician. Choice of the female physician is residualized with respect to differences in cost, travel time, physician experience, and patient ratings separately by specialty and respondent gender, with the subgroup mean added back. Points show mean adjusted choice probabilities within eight quantile bins of the female-minus-male wait-time difference, restricted to differences between -6 and 6 weeks. Solid lines show linear fits through the binned observations. The horizontal dashed line at 0.5 denotes equal probability of choosing the female and male physician. Vertical dashed lines indicate the implied indifference points, defined as the relative wait time at which the fitted probability of choosing the female physician equals 0.5.

Figure 6 shows the expected negative relationship between relative wait time and the probability of choosing the female physician: as the wait for a female physician becomes longer relative to the male physician, respondents become less likely to choose her. The implied indifference points, defined by where the fitted relationship crosses a choice probability of 0.5, differ substantially by respondent gender in urology and psychiatry. In urology, the implied indifference point is 1.4 additional weeks for women, compared with 0.4 weeks for men. In psychiatry, the corresponding values are 1.7 weeks for women and approximately zero for men.

In contrast, the fitted relationships in orthopedics show little evidence of a preference for gender concordance: the implied indifference point is 0.3 weeks for men and −1.0 weeks for women. These descriptive patterns therefore point to stronger gender-concordance preferences among women in urology and psychiatry, but not in orthopedics.

Willingness-to-Pay for Gender Concordance We next quantify these preferences using the mixed-logit estimates. Figure 7 plots the WTPs by gender and specialty. Pooling the three specialties, we find that women are willing to wait an extra 1.4 weeks to see a female physician, and this estimate is statistically significantly different from both zero and the male estimate, which is very close to zero.

Figure 7: Willingness-to-Pay in Wait Time for Gender Concordance



Notes: N = 2,793. Sample construction and estimation same as in Table 3. Bars represent point estimates and capped lines represent 95% confidence intervals. p-values result from the comparison of the male and female estimates within specialties.

In the gender-sensitive specialty, urology, women are willing to wait on average 1.6 additional weeks to see a female urologist. This estimate is statistically significant relative to no preference, and to the male estimate. Men do not exhibit a preference for urologists of either gender. A different pattern emerges in orthopedics, a specialty we chose for its relative neutrality on social and gender norms. Neither men nor women display a preference for a gender-concordant orthopedic surgeon, and the willingness-to-pay between the genders is not statistically distinguishable.

The female gender concordance preference is strong in psychiatry, a specialty we chose for its sensitivity to social norms (including gender), but lacking physical examination. Women

are willing to wait an average of 2.17 additional weeks to see a gender-concordant specialist. Men exhibit no statistically significant preference, and, if anything, the coefficient is negative. The difference between men and women for *concordant* physicians is statistically significant, but is not if we instead evaluate it through the lens of preference for *female* physicians.

To put these estimates in context, 1.6 and 2.2 weeks correspond to approximately 11 and 15 additional days. A 2025 survey found an average appointment wait of 31 days across six specialties in 15 major US metropolitan areas (AMN Healthcare, 2025), putting our estimates of willingness to pay an extra one-third to one-half of these average wait times.[8]

Beyond mean willingness-to-pay estimates, our mixed logit framework also estimates the degree of preference heterogeneity within each group. Panel D of Table 3 reports the estimated standard deviations of the random coefficients, which capture within-group preference heterogeneity not explained by observed characteristics. Two patterns are worth noting. First, the estimated standard deviation of the gender concordance coefficient is substantially larger for women than for men across all three specialties. For example, 0.621 for women versus 0.116 for men in urology, indicating that women's concordance preferences are considerably more dispersed around their group mean. This dispersion is consistent with the presence of distinct subgroups of women with very different underlying preferences, and motivates the heterogeneity analyses we present below. Second, the standard deviation of the wait time coefficient is small and similar across all groups (ranging from 0.074 to 0.118), indicating that sensitivity to wait time is approximately homogeneous across respondents. This is reassuring for the interpretation of our willingness-to-pay estimates: because the denominator of the willingness-to-pay ratio is precisely and homogeneously estimated, variation in willingness-to-pay for gender concordance across subgroups reflects true heterogeneity in concordance preferences rather than heterogeneity in the valuation of time.

4.3 The Role of Reported Prior Negative Healthcare Experiences Due to Gender

As we observed in Section 4.1, women are substantially more likely than men to report having been treated negatively or unfairly by a medical professional due to their gender. Those who report having had such experiences may place a higher value on gender concordance, anticipating that a female physician is less likely to perpetuate them. Alternatively, some other unobserved characteristic may drive both higher likelihood of perceiving mistreatment due to gender and higher gender concordance preferences.

We present results separately for women who report prior negative gender-related medical experiences and those who do not in Figure 8. Male respondents are excluded due to a very small share reporting such negative experiences. Pooling across all three specialties, women with prior perceived negative experiences are willing to wait more than twice as long for a gender-concordant physician than those without, approximately one additional week on average, although this difference is not statistically significant. This pattern holds across all three specialties, and with similar magnitudes. While the differences are not statistically

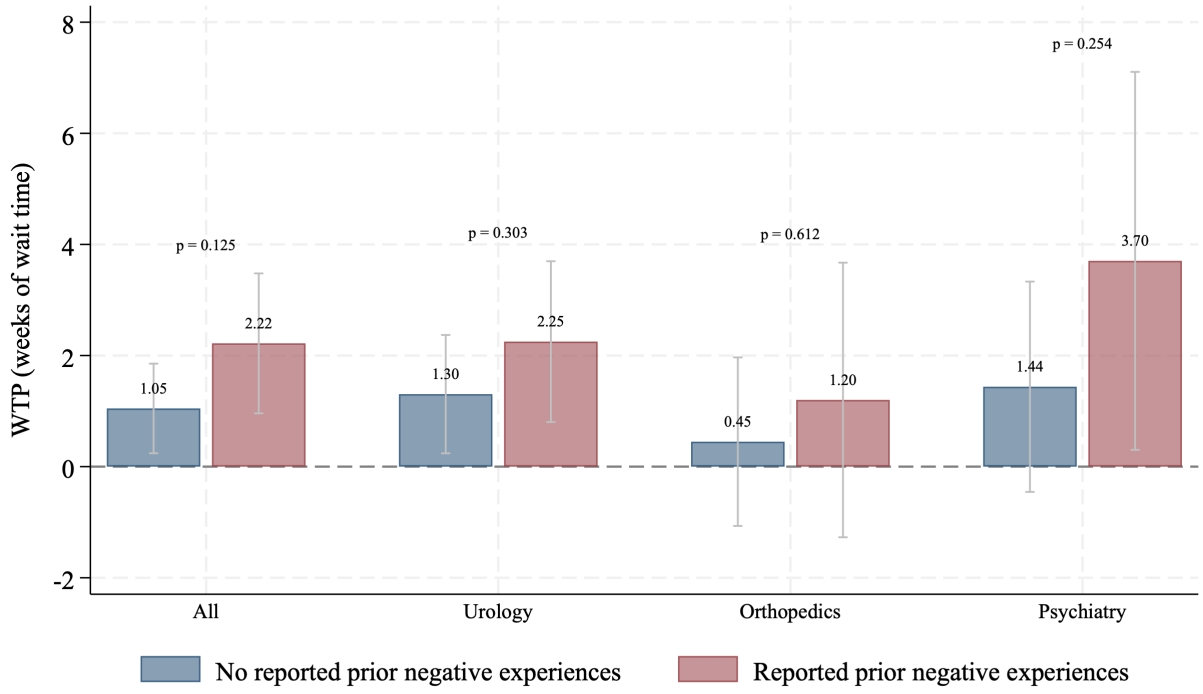
[8]The survey did not cover urology or psychiatry, so this figure can only provide a general sense of the magnitude.

significant, they all consistently point in the same direction: those women who report having had negative experiences in healthcare due to gender are willing to wait longer to see a female physician.

In urology, women who report prior negative experiences are willing to wait roughly 70% longer (2.25 weeks) as those without (1.3 weeks), and both estimates are statistically significantly different from zero. In orthopedics, the estimate for women who report prior negative experiences is almost three times larger, but neither group's estimate is statistically distinguishable from zero, consistent with the weak concordance preferences we observe in this neutral specialty overall. In psychiatry, women who report prior negative experiences are willing to wait an average of 3.7 additional weeks for a female psychiatrist, an estimate that is statistically significant, while the estimate for those without such experiences is 1.4 weeks and is statistically indistinguishable from zero. Psychiatry was chosen as a specialty because it may be sensitive to social norms, including gender, given patients discuss private and often sensitive topics with a psychiatrist. If negative experiences that are perceived to be based on gender fall into that category, this could explain the heterogeneous but strong preferences by past experience in this specialty.

These results suggest that gender concordance preferences are magnified by prior negative healthcare experiences. Women who report having encountered gender-based mistreatment in medical settings are willing to bear higher costs to see a female physician, and a substantial share (30%) of our female respondents do report such mistreatment. This has two implications. First, it suggests that concordance preferences partly reflect a response to past discrimination: women who believe their gender has negatively impacted past medical experiences are those most willing to wait for a female physician. Second, it implies that the welfare costs of a non-representative physician workforce fall disproportionately on those who may have already been negatively affected by it.

Figure 8: Female Willingness-to-Pay in Wait Time for Gender Concordance by Prior Negative Experiences



Notes: Sample construction and estimation same as in Table 3, limited to female respondents, who are further broken down by whether they reported having been treated negatively or unfairly by a medical professional due to their gender. Bars represent point estimates and capped lines represent 95% confidence intervals. P-values test the null hypothesis of no difference in gender concordance willingness-to-pay between those with and without reported negative experiences due to gender.

5. Real-World Application Using Medicare Data

The experimental results show that female patients prefer appointments with female physicians and are willing to wait longer to see them, particularly in specialties that are gender-sensitive and involve physical examinations, such as urology. As discussed in Section 2, female physicians are underrepresented in urology, with many areas of the country having no female urologist at all. If female patients are indeed waiting longer to access the limited supply of female urologists, the combination of gender concordance preferences and the under-supply of female urologists could carry meaningful real-world consequences for timely care.

Bladder cancer provides a useful setting to examine this question, and was also the condition motivating the hypothetical scenario in urology in our survey experiment. The main symptom of bladder cancer is blood in the urine, often accompanied by pain with urination and frequent urination (NHS, 2025). Because these symptoms overlap with more common conditions such as urinary tract infections, patients may not recognize the urgency and may delay care in favor of other appointment characteristics. While incidence is higher among men, women are diagnosed at later stages on average and face a worse prognosis

(Pignot, Barthélémy and Borchiellini, 2024). McDevitt and Roberts (2014) find that markets without a female urologist have higher mortality rates for female bladder cancer patients, all else equal. Our experimental results offer a mechanism: female patients reporting symptoms consistent with bladder cancer indicated a preference for waiting longer in order to see a female urologist. In this section, we show that patterns of realized patient-physician matching in Medicare claims are consistent with the experimental findings.

5.1 Summary Statistics in Medicare Claims

Data & Measurement We use Medicare Fee-for-Service (FFS) claims from 2006–2018, which comprise a random 20% sample of all Medicare FFS beneficiaries. Medicare is a useful setting to study bladder cancers because the disease disproportionately affects older adults, with roughly nine in ten diagnoses occurring in patients over age 55 (American Cancer Society, 2024). We focus on the carrier file, which includes all claims submitted by individual providers. Each claim includes the date of service, provider NPI, diagnosis codes, and procedure codes, allowing us to reconstruct the full sequence of outpatient encounters in the period leading up to a bladder cancer diagnosis. We restrict the sample to beneficiaries with a first observed diagnosis of bladder cancer starting at the age of 66.[9]

There is no direct measure of wait times in Medicare claims, and Medicare fee-for-service does not typically require a referral to book an appointment with a specialist. However, in practice, patients often present first to a primary care physician before scheduling with a urologist. We define wait time as the number of days between a patient's last primary care claim and their first urology contact, capturing the delay between initial symptom presentation and specialist evaluation.

We identify first urology contact working backward from the diagnosis date. Because the diagnosis date is not necessarily the date of first urology contact (many patients have related claims from laboratories and other providers in the days or weeks preceding a formal diagnosis), we identify the first urologist seen within the year prior to diagnosis as our anchor. In cases where no urology claim is observed before diagnosis but the diagnosing provider is a urologist, we treat that provider as the first urologist contact. This gives us a sample of 141,800 patients who were diagnosed with bladder cancer and saw a urologist within one year prior to diagnosis.

To identify the referring appointment in primary care, we combine two approaches; both are restricted to appointments within the year preceding the urology visit. First, where the referring physician's NPI is recorded and appears as the performing provider on a prior claim, we take the date of the most recent such visit. This identifies a referring appointment for 24.8% of patients. For the remainder, we use the patient's most recent primary care visit within the preceding year as a proxy for the referring appointment. We validate this proxy against the observed referral date in the subsample where both are available: the two dates match exactly 71.8% of the time. Among the 30.9% of non-matches, the proxy *under*states the wait in 76.7% of cases, typically because an unrelated primary care visit falls between the true referral and the urology visit, and overstates it in the rest, when the

[9]We exclude beneficiaries with a first observed diagnosis of bladder cancer at age 65, because these patients are more likely to have received a diagnosis before enrolling in Medicare insurance.

referral originates outside primary care. We apply the primary care measure to the 30.9% of patients for whom the referral is not directly observed but a qualifying primary care visit exists. The remaining 44.3% of patients have neither measure, likely meaning they reached urology without a referral, were referred by a non–primary care provider, or received prior care outside Medicare. These patients are excluded from wait-time analyses. While the inferred wait time measure is subject to measurement error, there is no reason to expect mismeasurement to differ systematically between male and female patients.

Summary Statistics Table 4 reports summary statistics for the 141,800 beneficiaries in our bladder cancer cohort, separately by patient gender. Women make up about 29% of the sample, consistent with the roughly three-to-one male-to-female incidence of bladder cancer. The average patient is 77 years old at diagnosis; women are slightly older (77.8 versus 77.2 years for men). Use of diagnostic tests is similar across genders: around 68% of patients have a urinalysis and 29% a urine cytology.

The largest differences arise in the care pathway. Women are more engaged with primary care than men - they have more primary care visit days (9.0 versus 7.7) and are less likely to have no primary care contact at all (7.5% versus 10.3%). Yet women are substantially less likely to reach a urologist. Only 74.5% of women have any urology visit, compared with 83.7% of men, and the gap is larger for a urology visit before diagnosis (62.7% versus 73.6%). Consistent with this, women are less likely to be diagnosed by a urologist (32.4% versus 37.6%) and less likely to undergo cystoscopy at their first urologist visit (31.9% versus 35.1%). Conditional on having a measured wait, women wait slightly less than men: 33.8 versus 37.1 days to see a urologist, and 38.8 versus 43.5 days to diagnosis.

Finally, female urologists are scarce: only 4.2% of patients' see a female urologist in their first visit. Women are more than three times more likely than men to see a female urologist (8.7% versus 2.6%), but gender concordance remains rare for women (8.7%) and nearly universal for men (97.4%).

These patterns carry several implications. First, the main gender disparities appear on the extensive margin (seeing a urologist at all), as well as on the intensive margin (conditional on seeing a urologist, seeing a gender-concordant one). Women are more likely to be in contact with primary care, yet less likely to be seen by a urologist, which points to the referral decision as a juncture at which women's care diverges from men's. This is consistent with clinical evidence that hematuria in women is more often incorrectly attributed to benign causes such as urinary tract infections (Garg et al., 2014), but could also be due to delays in care, due, for example, to strong preferences for gender concordance. Second, the shorter waits women experience do not equate to better access. Because women are screened out more heavily before reaching urology, those who are referred are plausibly the more urgent cases, which would mechanically shorten their observed waits. Comparisons of wait times conditional on reaching a urologist may therefore understate women's disadvantage. Third, the scarcity of female urologists means that even strong patient preferences for a female physician can be satisfied for a only small share of patients.

Table 4: Summary statistics, Medicare 20% sample

	All	Female	Male	p (F vs M)
N	141,800	40,417	100,540	
A. Demographics				
Age at diagnosis	77.4	77.8	77.2	<0.001
Age 65-69	0.190	0.189	0.189	0.844
Age 70-74	0.215	0.205	0.219	<0.001
Age 75-79	0.207	0.199	0.211	<0.001
Age 80-84	0.187	0.182	0.189	0.004
Age 85+	0.202	0.226	0.192	<0.001
B. Clinical				
Diagnosed by urologist	0.361	0.324	0.376	<0.001
Any urinalysis	0.682	0.689	0.681	0.003
Any urine cytology	0.289	0.294	0.288	0.019
Cystoscopy at first urologist	0.341	0.319	0.351	<0.001
C. Care pathway				
Any urology visit	0.809	0.745	0.837	<0.001
Urology visit before diagnosis	0.702	0.627	0.736	<0.001
No primary care visit	0.096	0.075	0.103	<0.001
Primary care visit days	8.0	9.0	7.7	<0.001
Unique urologists seen	1.0	0.9	1.1	<0.001
D. Wait for urology (days)				
Wait defined from observed referral	0.248	0.261	0.244	<0.001
Wait defined from last-PCP proxy	0.309	0.282	0.322	<0.001
No wait measure	0.443	0.457	0.435	<0.001
PCP→urology wait (days)	36.2	33.8	37.1	<0.001
PCP→diagnosis (days)	42.3	38.8	43.5	<0.001
E. First urologist				
Female urologist	0.042	0.087	0.026	<0.001
Gender-concordant urologist	0.740	0.087	0.974	<0.001

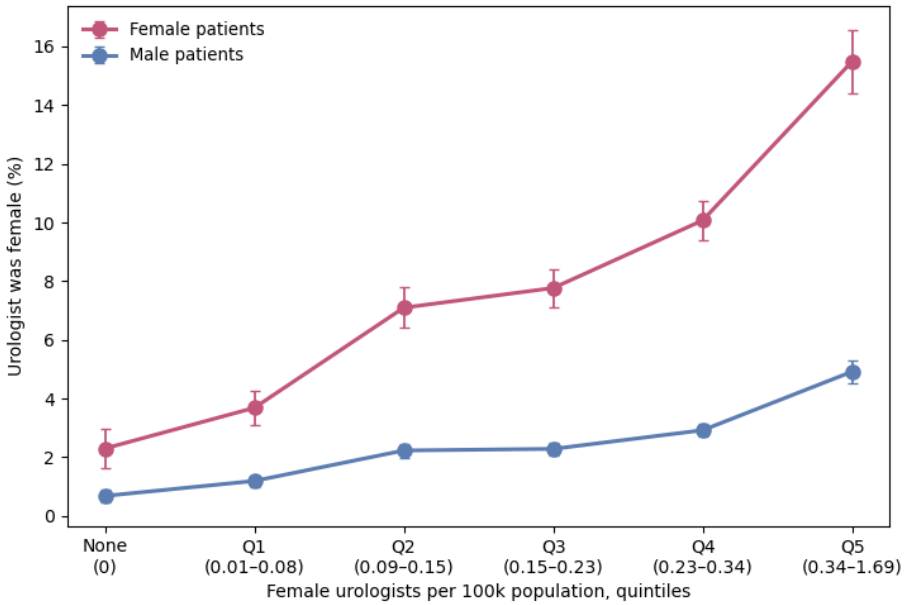
Notes: The sample consists of Medicare Fee-for-Service beneficiaries in the 20% sample with a first observed bladder-cancer diagnosis at age 66 or older and at least one urology contact within the year preceding diagnosis or a bladder-cancer diagnosis made by a urologist. All statistics are reported separately for female and male patients. Binary variables are reported as shares; continuous variables are reported as means. The final column reports p-values for tests of equality between female and male patients, using Welch's two-sample t-test for continuous variables and Pearson's chi-squared test for binary variables. "Diagnosed by urologist" indicates that the provider associated with the first observed bladder-cancer diagnosis was a urologist. "Urology visit before diagnosis" indicates an observed urology encounter prior to the diagnosis date. "Primary care visit days" is the number of distinct days with a primary care encounter, and "unique urologists seen" is the number of distinct urologists observed for the beneficiary. Wait time is measured as the number of days between the inferred referring encounter and the first urology contact. When the referring provider's NPI is observed and can be linked to a prior encounter, the most recent such encounter is used ("Wait defined from observed referral"). Otherwise, the most recent primary care visit within the preceding year is used as a proxy ("Wait defined from last-PCP proxy"). "No wait measure" identifies beneficiaries for whom neither measure is available. "PCP→diagnosis" is the number of days between the most recent qualifying primary care encounter and bladder-cancer diagnosis. "Female urologist" indicates that the first observed urologist was female. "Gender-concordant urologist" indicates that the patient and first observed urologist were of the same gender.

5.2 Wait times and realized gender concordance

Our primary question of interest is whether women with bladder cancer are waiting longer for urology appointments in order to see female urologists. We first estimate how patient-physician matching responds to the supply of female urologists. We identify urologists and their gender from Medicare Data on Provider Practice and Specialty (MD-PPAS). Because MD-PPAS does not record practice ZIP or HRR, we merge each urologist's NPI to the Carrier line file and assign the urologist to an HRR-year based on where they billed, taking each NPI's dominant HRR within a year. Aggregating to the HRR × year level gives the count of urologists, total and by gender, in each market and year. We merge these counts with the 2012 HRR-level population from the Dartmouth Atlas of Health Care to compute per-capita rates.

Figure 9 plots the share of patients matched to a female urologist against the local supply of female urologists per capita, by patient gender. For both male and female patients, the match rate with female physicians rises with supply, but the gradient is far steeper for female patients. This pattern is consistent with the gender-concordance preferences for urology documented among female respondents in the survey, and suggests that the limited supply of female urologists constrains the share of female patients who are able to match with one.

Figure 9: Share of Patients Matching with Female Urologists, By Patient Gender



Notes: Figure shows the share of patients whose urologist was female, by patient gender, across quintiles of female urologist supply in the patient's Hospital Referral Region (HRR). Supply is measured as the number of female urologists per 100,000 HRR residents, with urologist counts from MD-PPAS and HRR populations from Dartmouth Atlas. HRRs with no female urologist are shown separately; the remaining HRRs are grouped into quintiles, with bin cutoffs reported below the axis. If a patient has a claim with more than one urologist, we use the gender of the first urologist seen. Capped lines denote 95 percent confidence intervals for the share of patients seeing a female urologist.

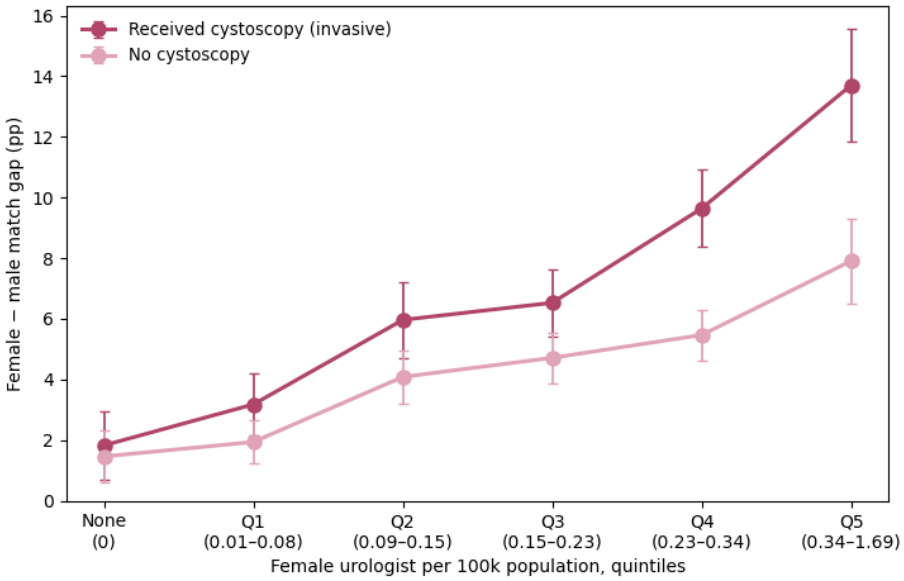
In our survey scenario, patients in the urology arm were told that during the hypothetical

appointment the doctor would perform a cystoscopy, a procedure in which a thin tube with a camera and light is passed through the urethra to examine the inside of the bladder. We described the procedure to mimic what a primary care physician might do, and to emphasize that the appointment would involve an intimate physical examination.[10] We would expect an anticipated intimate or invasive procedure to drive stronger preferences for gender concordance.

We examine whether this pattern appears in real appointments in the Medicare claims data by splitting patients by whether they received a cystoscopy, a standard procedure in the diagnosis of bladder cancer, during their first urology visit. In Figure 10 we plot the gap between female and male patients in the probability of matching with a female urologist, across quintiles of female urologists per capita in each patient's HRR. The darker pink line shows the gap for patients who received a cystoscopy at their first urology visit; the lighter pink line shows those who did not. The gap is positive in all quintiles, meaning female patients are more likely to be matched with female urologists, and it increases with supply of female urologists for both groups. The gradient is steeper for patients who received a cystoscopy, suggesting that those arranging appointments involving a more intimate procedure are more responsive to the supply of female physicians. This is consistent with the hypothesis that concordance preferences are stronger for women when the appointment involves a gender-sensitive examination. However, this figure is purely descriptive, as we do not observe what procedures patients anticipated when booking the visit, and the two groups may differ in other ways.

[10]The exact wording of the described appointment is shown in Figure B6.

Figure 10: Sorting into Female Urologists, by Cystoscopy Receipt



Notes: Figure plots the gap between female and male patients in the share matched to a female urologist (in percentage points). The x-axis shows quintile of female urologists per 100,000 HRR residents; HRRs with no female urologists are shown separately. If a patient has a claim with more than one urologist, we use the gender of the first urologist seen. The sample is split by whether the initial urology visit involved a cystoscopy. Capped lines denote 95 percent confidence intervals for the female–male gap.

We then examine wait times. If female urologists are in excess demand by female patients, then we would expect female patients waiting for a female urologist to wait longer. To test this we estimate the following equation separately for male and female patients. For patient i, in HRR h, in year t:

$$WaitTime_{iht} = \alpha + \beta \mathbb{1}[\text{Female urologist}]_i + \gamma_t + \mu_h + \delta X_{ht} + \varepsilon_{iht} \tag{9}$$

where $\mathbb{1}[\text{Female urologist}]_i$ indicates that the first urologist patient i saw was female, γ_t are year-of-diagnosis fixed effects, μ_h are HRR fixed effects, and X_{ht} is a control for the number of female urologists per 100,000 population of the HRR. $WaitTime_{iht}$ is the number of days between the last primary care visit and the first urology contact for patient i, in HRR h, in year t. Standard errors are clustered at the HRR level. Estimating the specification on male patients serves as a placebo. If female urologists practice in settings where wait times are longer for reasons unrelated to patient gender, the coefficient should be positive for men as well.

Results are shown in Table 5, Panel A. Female patients who see a female urologist wait 4.13 (12.2%) days longer than female patients who see a male urologist, against a mean wait of 33.8 days. Adding year and HRR fixed effects reduces it to 3.91 days (11.6%). For male patients, the corresponding estimates are small, negative, and statistically indistinguishable from zero in every specification. The null for male patients suggests the increased wait time to see a female urologist is specific to female patients matching on gender, rather than a general feature of female urologists' practices. Male patients face similar wait times regardless

of the gender of the urologist they see, suggesting they are not trading off wait time against physician gender in the way female patients appear to be.

To connect this wait-time result to the intimate-exam mechanism suggested by the survey and evidence on sorting by cystoscopy status, Table 5, Panel B interacts the female-urologist indicator with an indicator for whether the first urology visit involved a cystoscopy. Among women, the concordance wait penalty is concentrated in visits involving a cystoscopy. The estimated penalty is 4.74 days for these patients and is statistically significant, compared with 2.65 days for women whose visit did not involve a cystoscopy, which is not statistically distinguishable from zero. The interaction term is positive, consistent with stronger concordance-driven delay when the appointment involves an intimate procedure, but it is not itself precisely estimated, so we read this as suggestive rather than conclusive. Reassuringly, the male placebo shows no corresponding pattern in any specification. Combined with the steeper cystoscopy sorting gradient in Figure 10, the direction of these estimates is consistent with concordance preferences being strongest when care involves a gender-sensitive examination.

Table 5: Wait Times to See Female Urologists, by Patient Gender

	Wait time, PCP to urologist (days)					
	Female			Male (placebo)		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: All visits						
Female urologist	4.13**	4.27**	3.91**	-1.05	-1.14	-1.40
	(1.30)	(1.30)	(1.32)	(1.40)	(1.40)	(1.46)
Panel B: Interaction with cystoscopy						
Female urologist	3.16	3.28	2.65	1.20	1.06	0.95
	(2.03)	(2.03)	(2.04)	(2.32)	(2.33)	(2.34)
Female urologist × Cystoscopy	1.62	1.64	2.09	-3.70	-3.64	-3.96
	(2.80)	(2.80)	(2.80)	(2.75)	(2.75)	(2.68)
= Female urologist, cystoscopy visits	4.78**	4.91**	4.74**	-2.50	-2.58	-3.01
	(1.79)	(1.79)	(1.81)	(1.62)	(1.62)	(1.65)
Mean wait (days)	33.8	33.8	33.8	37.1	37.1	37.1
Year fixed effects		✓	✓		✓	✓
HRR fixed effects			✓			✓
Observations	21,892	21,892	21,892	56,707	56,707	56,707

OLS estimates. The outcome is days between the last primary care visit and first urology contact. *Female urologist* indicates the first urologist seen was female. Columns 1–3 are women; 4–6 repeat the specification on men as a placebo. In Panel B, *Cystoscopy* indicates the first urology visit involved a cystoscopy; the "= Female urologist, cystoscopy visits" row is the sum of the main effect and the interaction (the concordance wait effect among cystoscopy patients), with its SE computed from the full coefficient covariance. All columns control for female urologists per 100,000 population; Panel B additionally includes a cystoscopy main effect (coefficients not shown). Standard errors clustered by HRR in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

6. Real-World Application Using Online Patient Ratings

The concordance preferences documented in our survey also generate predictions about observed physician quality. Consider a specialty in which patients hold strong concordance preferences and the patient population is dominated by one gender: all else equal, physicians of the opposite gender will face lower demand. To attract patients, discordant physicians must compensate on other dimensions, such as perceived quality. We therefore would expect discordant physicians to be positively selected on these dimensions in markets where the majority of patients are discordant and where patients exhibit concordance preferences.

We examine if this is the case for two specialties that have these characteristics. The first is obstetrics and gynecology, where the patient population is female. Given the concordance preferences we document among female patients, particularly in gender-sensitive specialties, male obstetricians and gynecologists (OB/GYNs) will likely face lower demand at a given level of quality, and so we expect them to be positively selected and to receive higher ratings on average. The second is urology, where most, though not all, patients are male. The same mechanism predicts no positive selection for discordant preferences in urology. Male patients exhibit no concordance preference in our survey, so female urologists face no comparable demand penalty among the patients who make up most of their practice. In fact, we expect female patients in urology to have preferences for gender concordance. Thus, all else equal, even in urology, we may expect male physicians to be more positively selected if, despite women comprising a relatively small share of patients, concordance preferences increase demand for female physicians.

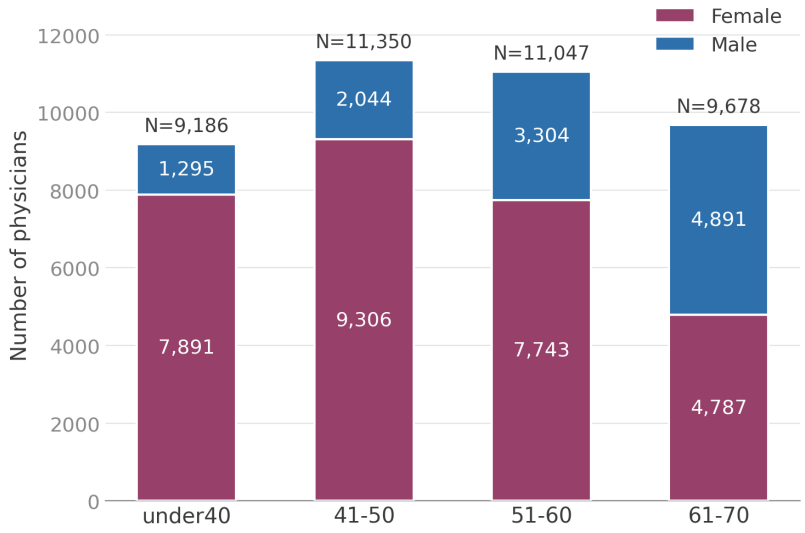
To test this prediction, we use patient reviews from the online platform Healthgrades. Healthgrades is a widely used platform for finding a doctor; the website reports that over half of all Americans who go to a doctor visit Healthgrades. Patients can use Healthgrades to find physicians, and in some cases directly book appointments. The site allows patients to filter search results by physician characteristics including gender, age, specialty, and insurance acceptance. Patients can also leave or view public reviews, which are listed at the bottom of physician profiles. Reviews are on a scale of one to five stars, with an optional text entry, and over 16 million physician reviews have been left by patients on Healthgrades in total.

Healthgrades lists 41,261 OB/GYNs,[11] of whom 79% have at least one patient rating. Of those with at least one rating, the average number of patient reviews is 18.68. Female OB/GYNs significantly outnumber male OB/GYNs. On Healthgrades, 72% of all OB/GYNs are female but this share varies with physician age (Figure 11), with older cohorts being more gender-balanced.

Figure 12 plots average patient ratings by physician age bin, separately for male and female OB/GYNs. Within every age bin, male OB/GYNs are rated significantly higher than female OB/GYNs. This is consistent with the preferences we estimate, and the dimension of variation (gender-sensitivity) that matters within each specialty. Facing a patient population with strong

[11] This is a high share, but likely not a complete count of OB/GYNs in the US. As of 2025 the national supply of OB/GYNs was estimated to be 49,170 full-time equivalent (FTE) (Silvestre and Lazenby, 2026). However, the average working hours for OB/GYNs is greater than 40 hours per week, suggesting the headcount of OB/GYNs is below 49,170.

Figure 11: Gender of OB/GYNs by Age Bin



Notes: Figure shows the count of OB/GYNs listed on Healthgrades, by physician age bin and gender.

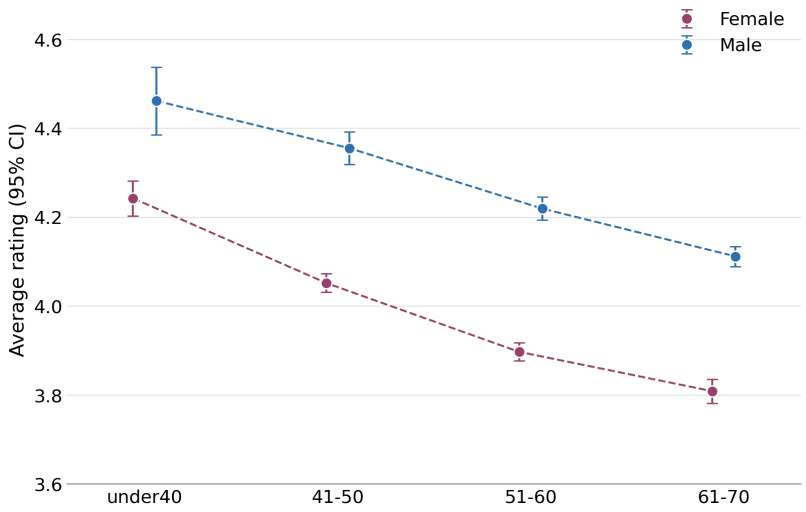
preferences for female physicians, male OB/GYNs must compensate on other dimensions (such as quality) to attract patients. This evidence is descriptive, as the sample of patients who leave a review is selected and other explanations are possible: male physicians may be perceived as higher quality across all specialties, or patients may hold lower expectations of male OB/GYNs and be more likely to leave a review after a positive interaction. Consistent with differential review behavior, 87.2% of male OB/GYNs have at least one review, compared with 75.7% of female OB/GYNs.

Urology provides a useful contrast: although the specialty is dominated by men, we do not expect a female premium in patient ratings. Estimates put the share of male urology patients between 60 and 70%, with considerable variation by condition (Ficko, Li and Hyams, 2018; Passarelli et al., 2025). Male urologists therefore have a large pool of patients indifferent to physician gender, and if anything, a small pool of patients that have an opposite preference. Therefore, despite constituting a majority of the specialty, we do not expect male urologists to face negative selection due to demand. We therefore predict no or positive rating premium for male urologists.

We also separately identify the subset of urologists who list male reproductive disorders and procedures as a practice area, which isolates physicians treating conditions specific to male patients. Practice areas on Healthgrades are self-reported, and urologists listing this area may also practice in other areas of urology and see female patients. However we would expect these specialists to attract a higher share of male patients. In this subspecialty, given the overwhelming share of male patients, we expect no difference in physician ratings by gender, given the lack of gender concordance preferences among men.

Opposite to OB/GYN, the urology workforce has a significantly higher share of male physicians. Overall, 81% of urologists are male, and 94% of urologists specialize in male reproductive disorders. Figure 13 shows the gender composition of urologists by age bin, for

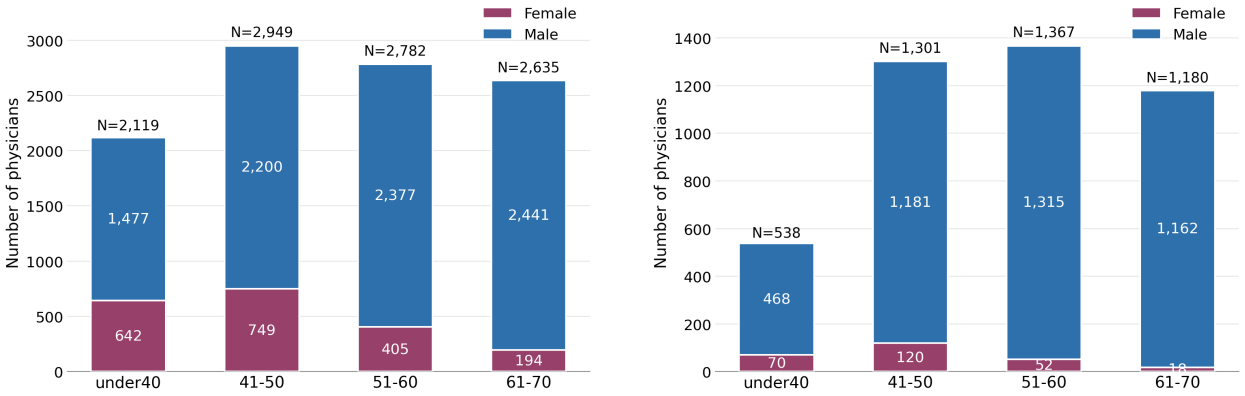
Figure 12: Average patient rating of OB/GYNs, by physician gender and age



Note: Figure shows average ratings of male and female OB/GYNs, by age bin. Patient ratings come from Healthgrades, all physicians with at least one rating are included.

all urologists in Panel (a) and the subset who specialize in male reproductive disorders in Panel (b). As with OB/GYN, urology has seen an increase in share of female physicians over time. Urologists also have more reviews on average compared to OB/GYNs: 96% of all urologists have a review on Healthgrades, with an average of 37.5 reviews each. Male urologists are more likely to have at least one review (96.5% versus 93.1%).

Figure 13: Gender of urologists by age bin



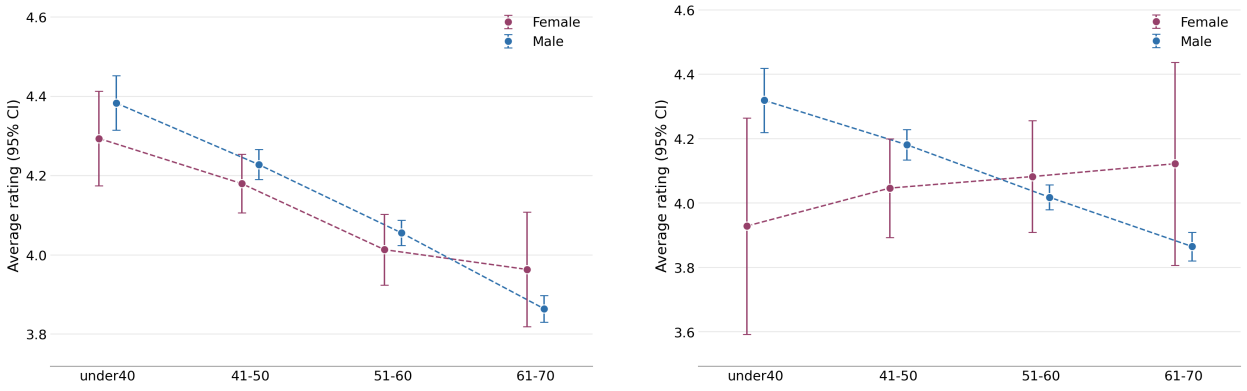
(a) All urologists

(b) Male-reproductive specialty

Figure 14 shows that, within age bins, there is no statistically significant difference in ratings between male and female urologists, although, as predicted, male urologists tend to have higher ratings across almost all age bins. The point estimates suggest slightly higher ratings for men in the younger bins and slightly higher ratings for women in the oldest bin, and this pattern is more pronounced among male reproductive specialists. The number

of female urologists listing this practice area is small, however, so the estimates in that subsample are imprecise. Taken together, the OB/GYN and urology results are consistent with the predictions of the survey. We expect a male ratings premium where the patient population holds strong preferences for female physicians, and no gap where it does not. Other explanations remain possible, since the ratings data are descriptive only and two specialties differ in the gender composition of their patients as well as in the strength of concordance preferences.

Figure 14: Average rating of urologists by age and gender



(a) All urologists

(b) Male-reproductive specialty

Notes: The figure plots average Healthgrades ratings by physician age bin and gender for urologists in panel (a), and a subset of urologists who specialize in male reproductive disorders in panel (b). Circles denote mean ratings within each age bin and gender cell. Error bars denote 95% confidence interval of the mean.

7. Discussion

Many medical specialties have large gender gaps in the physician workforce, and prior literature has documented that some patients prefer to see physicians of the same gender. Yet evidence on whether patients act on these preferences is limited. This is especially important given the increasing agency patients are exercising in their healthcare choices, often using online search and information tools. While in the past physicians may have recommended a particular specialist, or patients may have felt unable to ask for a specific demographic of physician, patients now increasingly have the option to choose a physician and an appointment themselves. Given the asymmetry of information in healthcare, they may do so without fully considering potential health consequences, particularly with respect to the progression of disease with delay.

This paper provides direct experimental evidence on the willingness to pay for gender-concordant physicians across specialties that differ in gender sensitivity, and links those estimates to real-world care-seeking behavior. Our results confirm that gender concordance preferences impact preferences over appointments, but they are concentrated among women and in gender-sensitive specialties. We find women are willing to wait nearly two additional weeks to see a female urologist, while men exhibit little to no concordance preference across

all three specialties studied. In orthopedics, a more gender-neutral specialty, neither men nor women display statistically significant concordance preferences. In psychiatry, where gender can play an important role, we find that women's estimate is large in magnitude and statistically significant. The strength of concordance preferences in the more gender-sensitive settings, and their absence in orthopedics, suggests that these preferences are not a general aversion to opposite-gender providers, but are instead tied to the intimate or sensitive nature of specific clinical interactions.

These preferences are heterogeneous in important ways. Preference for gender concordance is stronger among women who report having previously experienced negative or unfair treatment in medical settings due to their gender. Approximately 30% of female respondents report such experiences, and these women are willing to wait roughly two additional weeks for a female physician, relative to those who have not had the same experience.

Using Medicare claims data, we find suggestive evidence that these preferences matter for real-world access to care. We focus on a specialty with strong experimental concordance preferences, urology, and examine a subset of patients that matches our hypothetical choice scenario: those ultimately diagnosed with bladder cancer. Consistent with the experimental results, female bladder cancer patients are substantially more likely to see a female urologist than their male counterparts. As the supply of female urologists in a HRR increases, all patients are more likely to match with a female urologist, but the gradient is far steeper for female patients. These patterns suggest that women's ability to act on the concordance preference documented in our experiment is constrained by the local supply of female urologists. Among female patients, we find that those who see a female urologist wait on average four days longer than those who see a male urologist. Male patients show no corresponding gap, consistent with the delay reflecting excess demand for female urologists by female patients, rather than differences in how female physicians practice. Overall, the patterns in urology appointments within Medicare claims data suggest that preferences for gender-concordant physicians among women with bladder cancer are obtained at the cost of delays in care.

Concordance preferences among female patients in gender-sensitive specialties also have implications for the perception of physicians of different genders. All else equal, we would expect male physicians in gender-sensitive practice areas with majority-female patients to have lower demand. We therefore expect such male physicians to compensate by improving quality. Using patient review data from Healthgrades, we find patterns consistent with this prediction. Male OB/GYNs receive systematically higher patient ratings than female OB/GYNs, suggestive of positive selection. In urology, which is similarly gender-sensitive but where the majority of patients are male, we find no comparable gender gap in ratings.

The findings of the survey, in conjunction with the Medicare application, have meaningful policy implications. To the extent that delayed care worsens health outcomes, the under-representation of women in certain specialties is a public health concern. Policies aimed at diversifying the specialty workforce, or mitigating the concerns of female patients seeing male physicians in these areas (e.g., via chaperone policies), may have downstream benefits for patient health. Notably, we do not find strong concordance preferences in non-gender-sensitive specialties like orthopedics, suggesting that female patients do not have an across-the-board preference for female physicians. We also find that concordance preferences are in part driven by past negative experiences with the healthcare system, suggesting that interventions targeting the quality of patient-physician interactions could play an important role.

References

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

American Cancer Society. 2024. "Bladder Cancer Risk Factors."

AMN Healthcare. 2025. "2025 Survey of Physician Appointment Wait Times and Medicare and Medicaid Acceptance Rates." AMN Healthcare Physician Solutions Survey Report. Survey conducted January 15–February 24, 2025, across 15 major metropolitan areas and six medical specialties.

Andrew, Alison, and Abi Adams. 2025. "Revealed Beliefs and the Marriage Market Return to Education." *The Quarterly Journal of Economics*, 140(3): 2107–2162.

Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review*, 53(5): 941–973.

Chan, Alex. 2026. "Discrimination Against Doctors: A Field Experiment." Conditionally accepted, *American Economic Review*. Available at http://tinyurl.com/chandiscrimination.

Chen, Yiwei, and Stephanie Lee. 2024. "User-Generated Physician Ratings and Their Effects on Patients' Physician Choices: Evidence from Yelp." *Journal of Marketing*, 88(1): 77–96.

Ficko, Zita, Zhongze Li, and Elias S. Hyams. 2018. "Urology is a Sensitive Area: Assessing Patient Preferences for Male or Female Urologists." 5(2): 139–142.

Fong, Christina M., and Erzo F. P Luttmer. 2009. "What Determines Giving to Hurricane Katrina Victims? Experimental Evidence on Racial Group Loyalty." *American Economic Journal: Applied Economics*, 1(2): 64–87.

Gallen, Yana, and Melanie Wasserman. 2023. "Does information affect homophily?" *Journal of Public Economics*, 222: 104876.

Garg, Tullika, Laura C. Pinheiro, Coral L. Atoria, S. Machele Donat, Joel S. Weissman, Harry W. Herr, and Elena B. Elkin. 2014. "Gender Disparities in Hematuria Evaluation and Bladder Cancer Diagnosis: A Population Based Analysis." *Journal of Urology*, 192(4): 1072–1077.

Greenwood, Brad N., Seth Carnahan, and Laura Huang. 2018. "Patient–Physician Gender Concordance and Increased Mortality among Female Heart Attack Patients." *Proceedings of the National Academy of Sciences*, 115(34): 8569–8574.

Lamiraud, Karine, Morgane Le Guern, Michael Rockinger, and Christine Sevilla-Dedieu. 2026. "Preference for Gender Concordant GP and Medical Deserts." *Review of Economics of the Household*.

McCarthy, Ian, Kaylyn R. Sanbower, and Leonardo Sánchez Aragón. 2024. "Online Reviews and Hospital Choice." Working paper.

McDevitt, Ryan, and James W. Roberts. 2014. "Market Structure and Gender Disparity in Health Care: Preferences, Competition, and Quality of Care." *The RAND Journal of Economics*, 45(1): 68–93.

Miyawaki, Atsushi, Anupam B. Jena, Lisa S. Rotenstein, and Yusuke Tsugawa. 2024. "Comparison of Hospital Mortality and Readmission Rates by Physician and Patient Sex." *Annals of Internal Medicine*, 177(5): 586–594.

National Cancer Institute. 2025. "Health Information National Trends Survey (HINTS): Public-Use Datasets, HINTS 3 (2008) through HINTS 7 (2024)." *https://hints.cancer.gov/data/download-data.aspx*, Bethesda, MD: National Cancer Institute, Division of Cancer Control and Population Sciences. Cycles used: HINTS 3 (2008), HINTS 4 Cycles 1 and 3 (2011, 2013), HINTS-FDA Cycles 1 and 2 (2015, 2017), HINTS 5 Cycles 1, 3 and 4 (2017, 2019, 2020), HINTS 6 (2022), HINTS 7 (2024). Accessed September 21, 2026.

NHS. 2025. "Symptoms of Bladder Cancer."

Nolen, Haley A., Justin Xavier Moore, Joel B. Rodgers, Henry E. Wang, and Lauren A. Walter. 2016. "Patient Preference for Physician Gender in the Emergency Department." *The Yale Journal of Biology and Medicine*, 89(2): 131–42.

Passarelli, Natalie, Alexa Steckler, Benjamin Homer, Alexander Homer, Simone Thavaseelan, and Elias Hyams. 2025. "Are Women Urologists Caring for Predominantly Women Patient Populations? Exploring the Role of Patient-Physician Gender Concordance in Robotic and Laparoscopic Urological Surgery." 201: 174–178.

Pignot, Géraldine, Philippe Barthélémy, and Delphine Borchiellini. 2024. "Sex Disparities in Bladder Cancer Diagnosis and Treatment." *Cancers*, 16(23): 4100.

Reyes, Jessica Wolpaw. 2006. "Do Female Physicians Capture Their Scarcity Value? The Case of OB/GYNs." National Bureau of Economic Research Working Paper 12528.

Reyes, Jessica Wolpaw. 2008. "Gender Preference and Equilibrium in the Imperfectly Competitive Market for Physician Services." *Eastern Economic Journal*, 34(4): 431–448.

Silvestre, Jason, and Gweneth B. Lazenby. 2026. "Geographic Variation in Supply, Demand, and Adequacy of the Obstetrics and Gynecology Physician Workforce: Forecasts and Shortage Risks in the United States." *Archives of Gynecology and Obstetrics*, 313(1): 95.

Tsugawa, Yusuke, Anupam B. Jena, José F. Figueroa, E. John Orav, David M. Blumenthal, and Ashish K. Jha. 2017. "Comparison of Hospital Mortality and Readmission Rates for Medicare Patients Treated by Male vs Female Physicians." *JAMA Internal Medicine*, 177(2): 206–213.

Varadarajulu, Shyam, Carol Petruff, and William H. Ramsey. 2002. "Patient Preferences for Gender of Endoscopists." *Gastrointestinal Endoscopy*, 56(2): 170–173.

Walker, Brigham, Janna Wisniewski, Sarah Tinkler, Jillian Torres, and Rajiv Sharma. 2024. "Identity and Access: Gender-Based Preferences and Physician Availability in Primary Care." *Journal of Economic Behavior & Organization*, 225: 707–722.

Wallis, Christopher J.D., Angela Jerath, Natalie Coburn, Zachary Klaassen, Amy N. Luckenbaugh, David E. Magee, Amanda E. Hird, Kate Armstrong, Damian Broe, Bhavya Ravi, Girish Kulkarni, and Raj Satkunasivam. 2022. "Association of Surgeon–Patient Sex Concordance With Postoperative Outcomes." *JAMA Surgery*, 157(2): 146–156.

Werner, Rachel M., Edward C. Norton, R. Tamara Konetzka, and Daniel Polsky. 2012. "Do Consumers Respond to Publicly Reported Quality Information? Evidence from Nursing Homes." *Journal of Health Economics*, 31(1): 50–61.

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

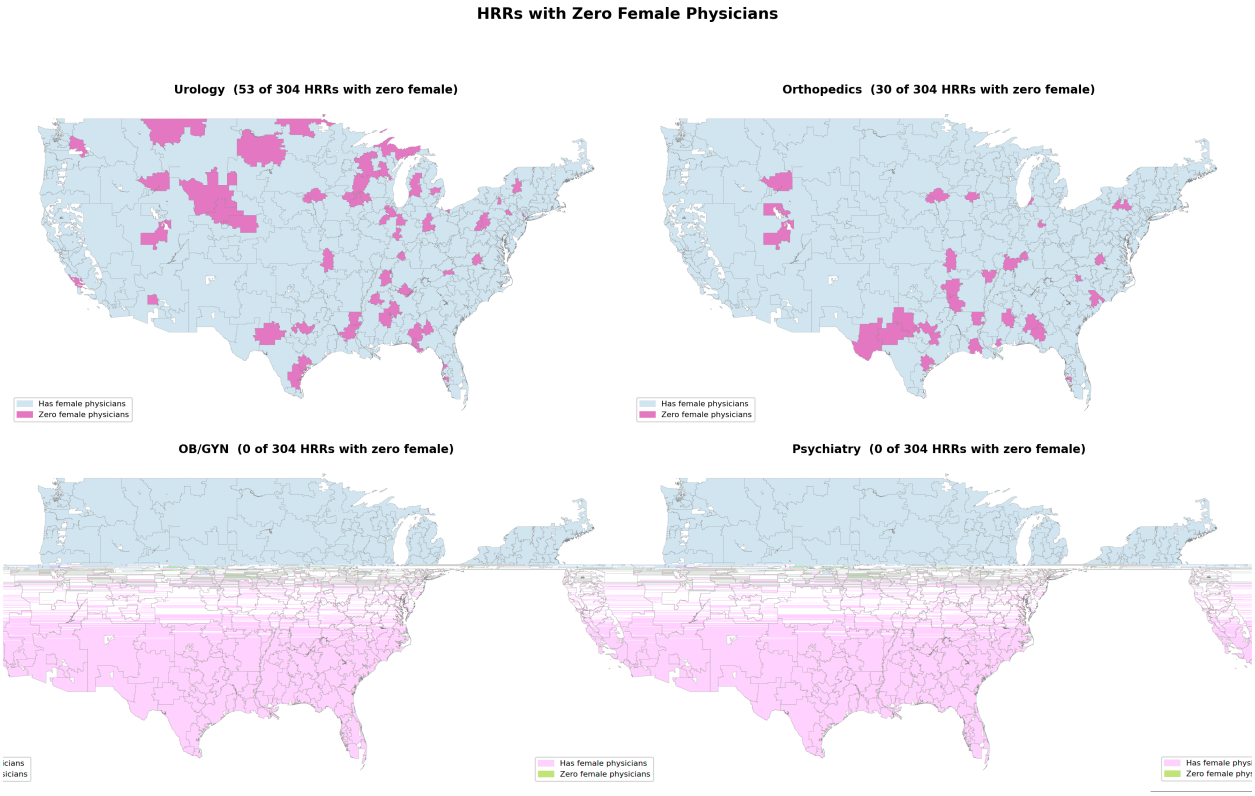
A. Background

Table A1: NPPES Physician Sample Construction

Step	N remaining	N dropped	% of total
Total NPI records	9,129,558	0	100.0%
Restrict to individuals (Entity Type = 1)	6,962,229	2,167,329	76.3%
Drop deactivated NPIs	6,947,618	14,611	76.1%
Restrict to physician taxonomy codes	1,204,934	5,742,684	13.2%
Require non-missing gender (M/F)	1,204,604	330	13.2%
Match to HRR via practice ZIP code	1,189,697	14,907	13.0%

Notes: NPPES NPI file for September 2025, downloaded from the NBER.

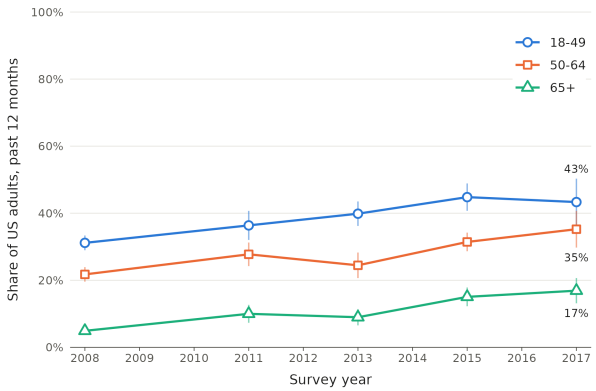
Figure A1: Hospital Referral Regions with Zero Female Physicians by Specialty



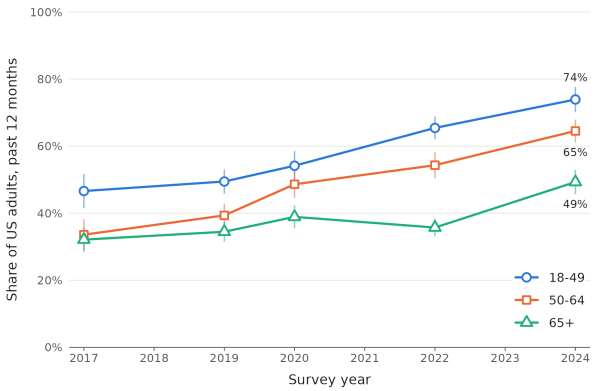
Notes: Data from NPPES NPI file September 2025, downloaded from NBER. Analysis restricted to top twenty specialties by number of physicians. Specialty inferred from provider taxonomy codes. Practice location linked to HRR via ZIP. 304 HRRs displayed (Hawaii and Alaska excluded).

Figure A2: Use of online tools to find and book providers, by age group

(a) Looked for a health care provider online



(b) Made an appointment with a provider online



Notes: Survey-weighted share of adults in each age group (18–49, 50–64, 65 and over) who did each activity in the past 12 months. The denominator is all adults in that age group, including those who do not use the internet. Vertical bars are 95% confidence intervals from 50 jackknife replicate weights. In panel (a), the 2017 point is from the HINTS-FDA cycle, which keeps the 2008–2015 "Internet" wording. In panel (b), the 2017–2020 item asks about "a computer, smartphone, or other electronic means"; the 2022–2024 item asks about "the Internet"; the dip among adults 65 and over in 2022 coincides with this change. Source: HINTS public-use files.

B. Detail on Survey Design

In this appendix, we outline details in our survey design and the motivation behind them.

B.1 Explanation of Hypothetical Choice

After consenting to the survey, passing an instructional manipulation check, and answering some background questions, respondents are introduced to the symptom scenarios and hypothetical choice component of the survey. We begin by showing them the following text.

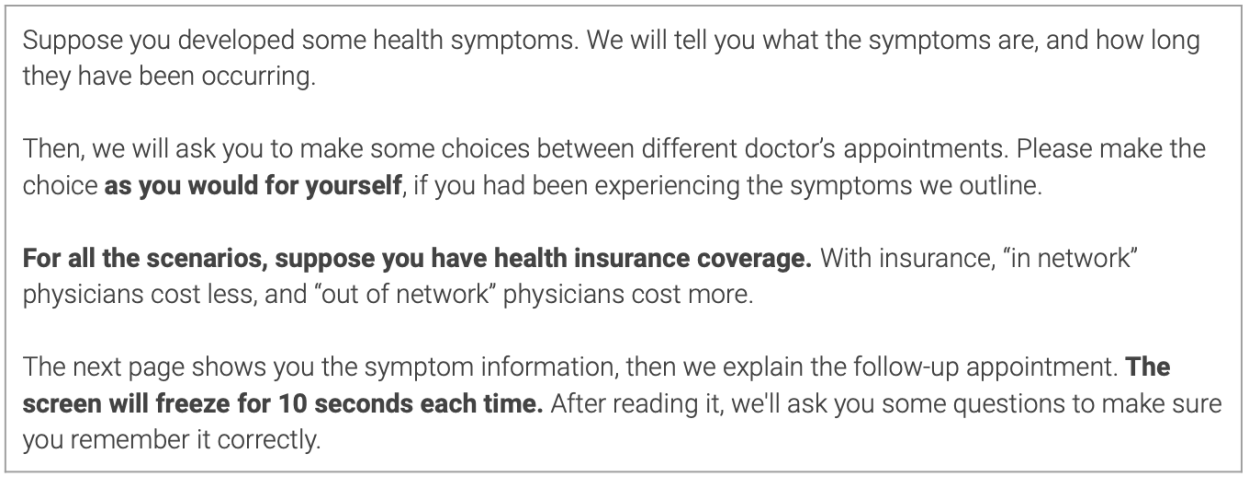


Figure B3: Introduction to Symptom Scenario and Hypothetical Choice

We made several important choices in this introduction:

- We emphasize that the hypothetical choices should be made from the perspective of the respondent, as if he/she were making the choice in reality ("Please make the choice **as you would for yourself**...").
- We asked respondents to imagine they had health insurance coverage when making their choice. This is because we wanted to include a monetary cost for each physician-appointment, to allow for calculation of a willingness-to-pay in dollar terms for concordance. Critically, we did not want this cost to be inferred as a measure of physician quality (e.g., more expensive physicians are better), so sought an explanation that was independent on underlying physician characteristics. We described physicians as in- or out-of-network to motivate differences in cost.
- We told respondents what information they would see, and that each information screen would be frozen for 10 seconds. Freezing screens can increase attrition in surveys, so we kept the time interval short and forewarned respondents to avoid confusion and motivate engagement with the material.

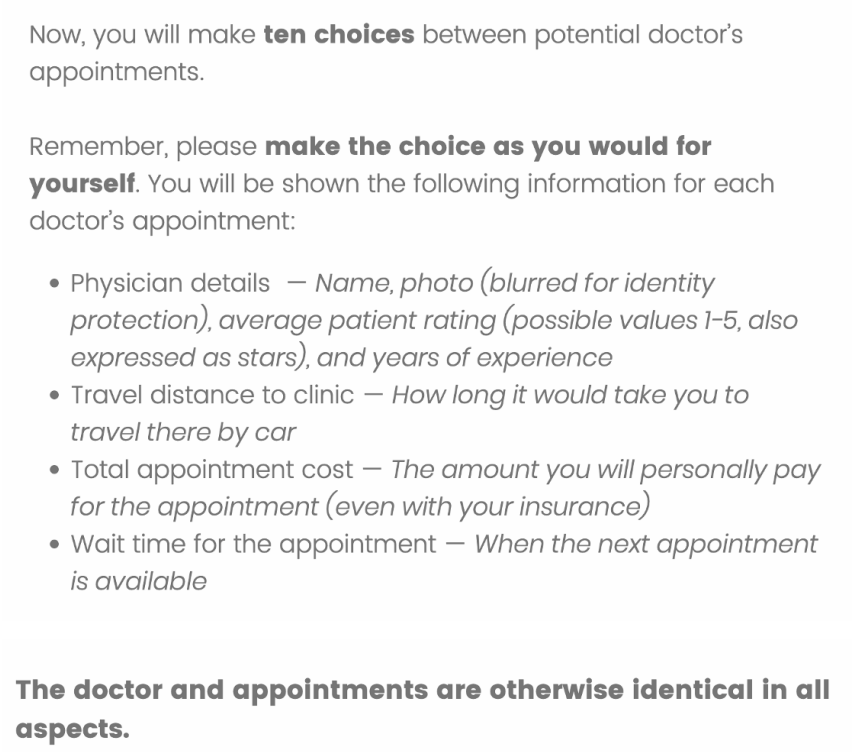


Figure B4: Preview to Hypothetical Choices

B.2 Information Treatments

After previewing the scenarios and hypothetical choice exercise, respondents were randomized into a speciality/symptom scenario. First, they were asked to suppose that they had been experiencing a collection of symptoms. We outline those symptoms and specialty/disease areas in Figure B5.

Neutral *Orthopedics* *Osteoarthritis Symptoms*	Over the past month, suppose you developed these symptoms: • Stiffness in your knee when getting up from your desk • A "scraping" sensation in your knee when you walk • A dull ache in your knee after walking for longer periods
Gender *Urology* *Bladder Cancer Symptoms*	Over the past month, suppose you developed these symptoms: • You have soreness in your hips and lower back • You need to wake up several times at night to urinate • You have noticed some blood in your urine on occasion
Social/Behavioral *Psychiatry* *Depressive Symptoms*	Over the past month, suppose you developed these symptoms: • You have been struggling to get to sleep at night • You do not feel as interested in your regular leisure activities • You have found it difficult to concentrate when at work • You have been feeling sad most days

Figure B5: Symptom Scenarios

We made several important choices in these symptom outlines:

- We kept the information load similar across scenarios. In all cases we said symptom development had been occurring for a month and detailed 3-4 symptoms.
- We tried to maintain a non-emotive tone with no symptoms that presented an obvious emergency that might warrant a visit to the emergency department. For example, in the case of depressive symptoms, we avoided any mention of self-harm. This was because we wanted the choice of organizing a specialist appointment, potentially several weeks in advance, to be a plausible one.
- We chose gender-neutral symptoms such that male and female respondents could experience the same symptom scenarios, to allow for a cleaner comparisons of male versus female preferences for concordance.
- We detailed symptoms found on reputable websites for specific diseases, such as the Centers for Disease Control and Prevention (CDC), Mayo Clinic, and National Health Service (of the United Kingdom).

Respondents were then told that they had been referred to see a specialist by a primary care physician. We made this choice to make the choice of seeing a specialist more plausible, and prevent respondents from reneging on the choice pathway we had laid out. For example, if we had asked them to choose between specialists, a respondent might have chosen visiting their usual primary care physician as an outside option. We then detailed exactly what type of specialist they had been referred to, as well as what would happen at the appointment. We outline this content in Appendix Figure B6.

Neutral *Orthopedics* *Osteoarthritis Symptoms*	A primary care doctor referred you to see an **orthopaedic surgeon**. An orthopaedic surgeon is a doctor who diagnoses and treats conditions related to the bones, joints, and muscles. At the appointment, the doctor will examine your knee, ask you about your symptoms, and might order some imaging tests (e.g. X-ray, ultrasound) and lab tests (e.g. blood tests). The doctor will also discuss possible diagnoses and treatment options.
Gender *Urology* *Bladder Cancer Symptoms*	A primary care doctor has referred you to see a **urologist**. A urologist is a doctor who diagnoses and treats conditions related to the urinary tract and reproductive system. At the appointment, the doctor will examine your abdomen and ask about your symptoms. The doctor will also examine the inside of your bladder by passing a thin tube with a camera and light at the end through your urethra (the tube through which you urinate).
Social/Behavioral *Psychiatry* *Depressive Symptoms*	A primary care doctor has referred you to see a **psychiatrist**. A psychiatrist is a doctor who diagnoses and treats conditions related to mental health. At the appointment, your psychiatrist will ask you about you and your family's medical history, your daily life, and your mental health. The appointment will be 1-1.5 hours long, to give you time to discuss your feelings in depth.

Figure B6: Specialist Scenarios

We made several important choices in these symptom outlines:

- We kept the format and length of the specialist introduction similar across treatments, to facilitate comparisons across the specialties.
- We detailed exactly what would happen at each appointment, in case the respondent did not know of the specialty or what typically occurred at such visits. Providing this information was also essential for realism, given it is in the realm of what a primary care physician would tell a patient in an initial visit.
- We highlighted elements of the appointment where the norms we were interested in came into play. So, in urology we explicitly stated that the appointment would include an intimate physical exam, given we are interested in how preferences for concordance might vary across specialties with different levels of physical engagement and gender norms. In psychiatry, we explicitly stated that the appointment would be 1-1.5 hours long, and would involve a detail review of the respondent's personal life.

After these information treatments, the respondents were asked to answer several questions about the information. As we detailed above, the respondents had been forewarned these questions would occur, as an incentive for reading the information carefully. Though their answers did not impact their compensation at all, answering the questions incorrectly extended the survey time; those who did not correctly answer all the questions were routed to the information for a second time.

The last section before the hypothetical choice asked respondents to react to the information they had been shown. These questions were designed to help us understand how

medical knowledge might impact willingness-to-pay for concordance with, for example, time to appointment. For example, we asked respondents how urgent they thought the issue was, as well as if they had any predicted diagnoses.

B.3 Examples of Physician Search Tools

Below we provide some examples of physician search tools that we drew on while designing our hypothetical choice experiment.

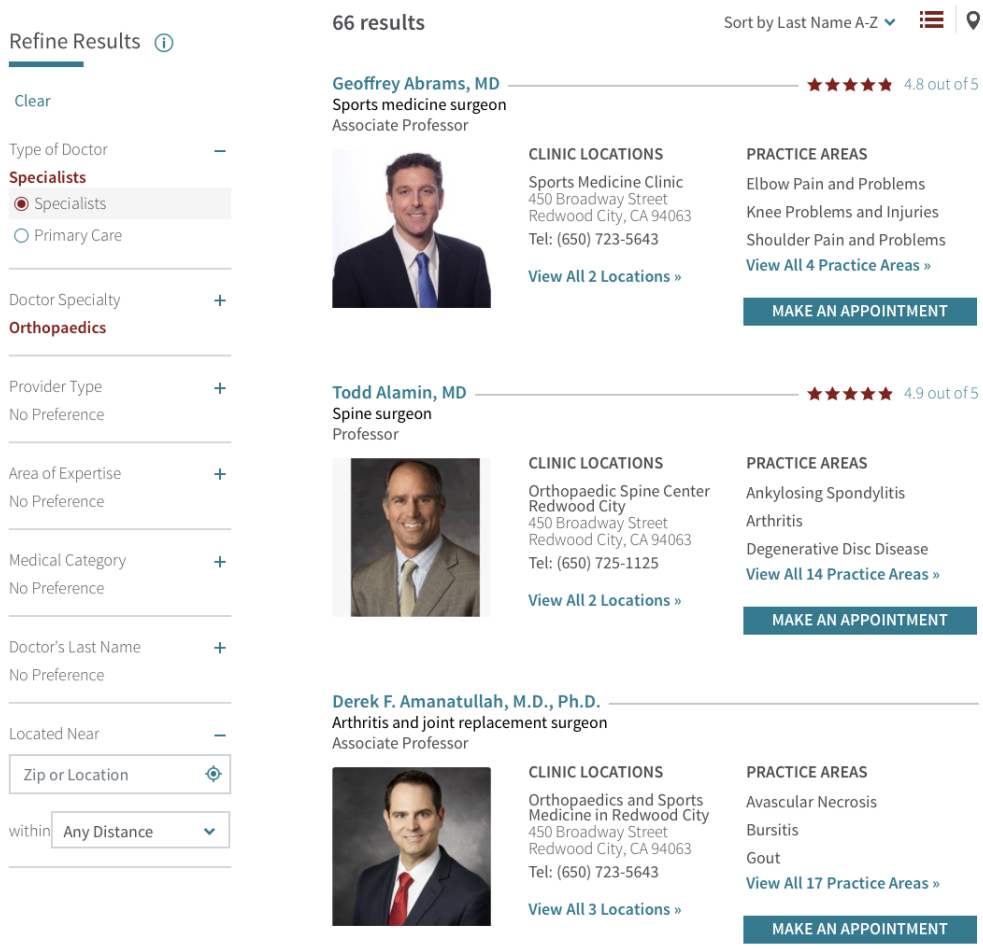


Figure B7: Search results for Stanford Orthopedics Specialists

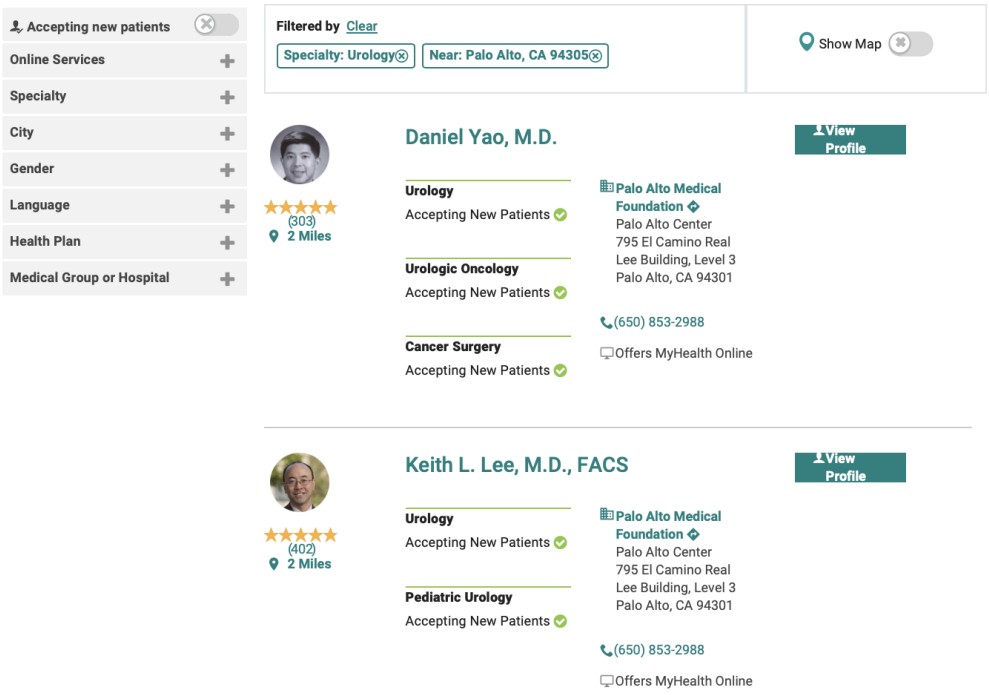


Figure B8: Search results for Sutter Health Urologists

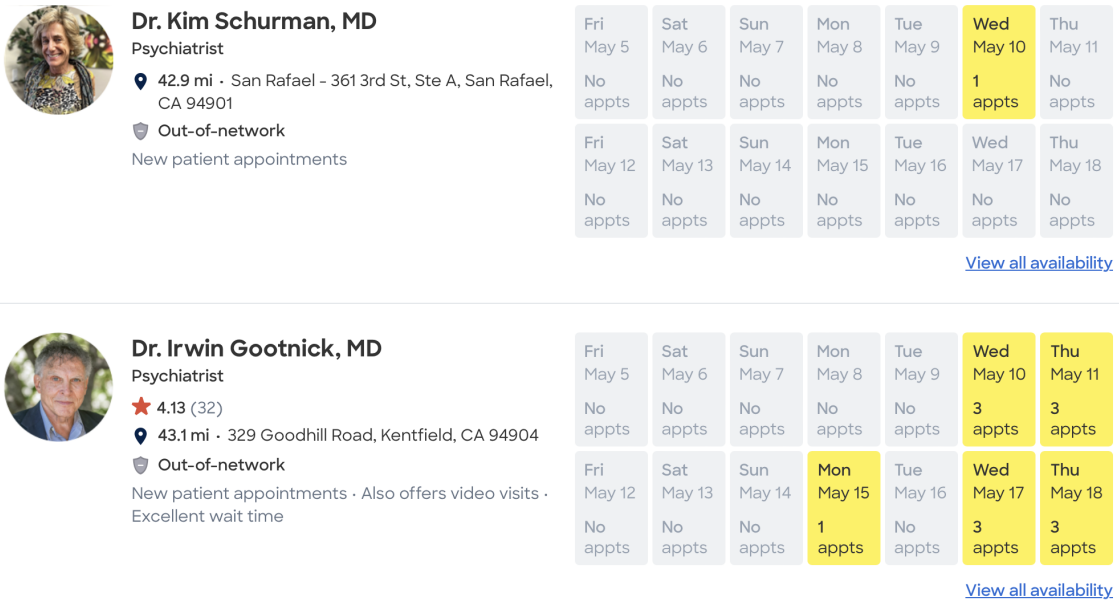


Figure B9: Search results for Zocdoc Psychiatrists

C. Experimental Results

C.1 Experimental Checks

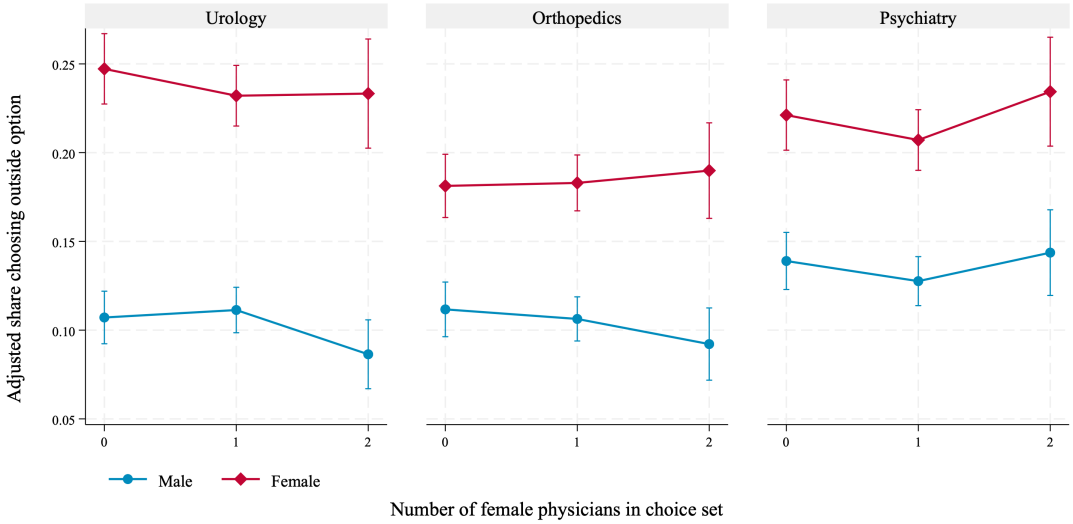


Figure C10: Outside-Option Choice by Physician Gender Composition

Notes: The figure plots the adjusted share of respondents choosing the outside option by the number of female physicians in the choice set, separately by specialty and respondent gender. Choice of the outside option is residualized with respect to the mean wait time, cost, travel time, physician experience, and patient rating of the two physician alternatives, separately by specialty and respondent gender, with the subgroup mean added back. Points show mean adjusted outside-option choice for choice sets containing zero, one, or two female physicians. Vertical bars denote 95% confidence intervals.