# Gender and the Market for Expertise: What can we Learn from Second Opinions?

*Draft: Not for Broader Distribution*

Mathijs de Vaan and Toby Stuart

**Abstract**

Despite the knowledge gap between a non-expert client and an expert, clients may be uncertain about the quality of the advice given by the expert. To reduce this perceived uncertainty, clients often rely on characteristics of the expert and the relationship with expert to decide whether to follow the advice provided. We argue that two such characteristics, status and trust, in the expert-client interaction introduces a gender gap in the extent to which advice is adopted and a subsequent gender gap in expert earnings. We examine the referrals of patients to medical specialists and the extent to which the patients see another specialist for a second opinion. These instances in which the advice of the first specialist is not directly adopted has the potential to shift earnings from the first specialist to the second. We document that female specialist are substantially more likely to see their patients visit another specialist for a second opinion. This difference is driven by the fact that 1) female patients sort into female specialists and also exhibit higher unconditional second opinion rates and 2) male patients strongly relying on specialist gender to guide their decision to request a second opinion. We discuss the implications of these findings for markets for expertise and link them to prior work that shows that men and women tend to hold women to the same higher standard. We then show that these patterns increase the gender pay gap between male and female specialists.

# Introduction

Prior research provides consistent evidence of gender differences in compensation and career advancement across the professions (Castilla, 2008; Pedulla and Thébaud, 2015; Quadlin, 2018). These differences exist and persist against a backdrop of shifting occupational choices and structures leading to a decades-long uptick in the proportion of female graduates in law, medical, and business schools (Blau and Kahn, 2013, 2017; Mann and DiPrete, 2013; Adams, 2010). Despite near gender equality in rates of entry into the advice-based professions, earnings disparities continue to the present (Beckman and Phillips, 2005; Azmat and Ferrer, 2017; Zeltzer, 2020; Gallotti and De Domenico, 2019).

We argue that the gender gap in earnings is partly rooted in client-side, gender-based differences in the perception of value of expertise in expert-client dyads. As contemporary scholars of the sociology of professions observe, expertise is transacted for in relational contexts that are cocooned in core, social processes (Azocar and Ferree, 2015). In the client-based professions, the expert role entails a situational diagnosis drawn from the knowledge base of a field, paired to recommendation(s) for treatment. Abbott (2014) describes these as mediating acts, in which diagnosis is pattern recognition against a professional knowledge system and treatment involves extracting an instruction set from it. Despite years of rigorous training that creates a facade of objectivity and certainty, the diagnosis-treatment sequence between expert and client is embedded in specific social relationships and cultural understandings. The ritualistic accoutrement orchestrated to create a veil of certainty surrounding these interactions belies frequent ambiguity in how to solve a client's problem, and what constitutes superior guidance (Fox, 1957; Eddy, 1984). This imbues the market for expertise with uncertainty.

Research has shown that in contexts characterized by uncertainty about quality, people rely on status markers in an attempt to overcome the perception of uncertainty (Podolny, 2010; Stuart et al., 1999; Ridgeway and Correll, 2004). And if status and quality are only loosely coupled, it is likely that resources are allocated to those in the top of the status hierarchy, not to those who provide the best service or advice. Although prior work has identified a host of salient status markers such as endorsements of prominent actors (Stuart et al., 1999), racial identity (Melamed et al., 2019), and occupational prestige (Freeland and Hoey, 2018), few status markers have the potential to be as consequential as gender. As Ridgeway and Correll (2004) suggests, shared understandings about gender are relatively homogeneous and expectations about performance differences between men and women are deeply ingrained in societal beliefs, which makes gender a status marker that is likely to shift resources away from women and towards men. Indeed, empirical research has demonstrated such a process in contexts including investing (Brooks et al., 2014) and student evaluations (Moss-Racusin et al., 2012). In the context of professional advice, prior work on gender as a status marker would suggest that male professionals may benefit from their higher status through higher demand for their services and lower rates at which advice is questioned.

Uncertainty, however, may not only be overcome through the use of status markers, but also through trust. Podolny (1994), for example, shows that when faced with uncertainty, actors increase their exchanges with partners whom they trust. In contexts characterized by the potential for repeated exchanges, trust may emerge over time (Gulati and Sytch, 2008), but in settings where repetition of exchanges is less likely or where actors typically have no history of social exchange, trust may emerge from other sources. One such source is similarity in salient characteristics (Ruef et al., 2003), with gender similarity being a prominent feature of the interaction (Kanter, 2008). In the context of advice seeking, men may trust the male expert more than the female expert and the female client may trust the female expert more than the male expert. While this finding contradicts research that shows that both men and women penalize female providers (Moss-Racusin et al., 2012; Brooks et al., 2014), it is consistent with research showing that many types of social interaction are characterized by a preference for gender similarity(Kossinets and Watts, 2009).

We analyze gendered patterns in clients' acceptance of expert advice by studying the frequency of second opinions obtained by patients that visit specialist providers in healthcare. Physician visits are an ideal context to examine these questions because we can observe population-level data and nearly every expert-client pair that occurs, with clear indicators of the gender of clients (patients) and experts (medical specialists) and with many details about the nature of the consultation. In a dataset comprising millions of physician visits in Massachusetts between 2010 and 2015, we first show that the probability of obtaining a second opinion varies with the gender of the patient. We find that *female patients*, on average, seek second opinions at a higher rate than do male patients. However, the likelihood of securing a second opinion conditional on seeing a *female specialist* is much larger for male patients than for female patients. In other words, compared to the decision of a female patient to obtain a second opinion, the decision of a male patient to do so is strongly guided by the gender of the specialist. Women seek second opinions at unconditionally higher rates, but male patients seeing female specialists are the most likely of all pair types to consult an additional physician. These patterns persist across a range of robustness checks.

Next, we examine the probability of switching to a specialist of the opposite gender, conditional on pursuing a second opinion. We find a similar pattern here: when the first specialist a patient sees is the opposite gender, male patients are *substantially* more likely to switch to a same-gender specialist in a second opinion visit, than are female patients. Finally, we conduct a back-of-the-envelope calculation to assess the implications of these patterns on earnings for specialists. We estimate that the missed earnings for female physicians as a fraction of their total earnings are roughly 15% higher than the missed earnings for male physicians because their expertise is questioned.

# Theory

When encountering problems or decisions that require specialized knowledge, clients consult experts to select a course of action. For example, companies hire bankers for financial advice (Bloom et al., 2020), clients seek counsel from attorneys in legal matters (Somaya et al., 2007), and patients confer with physicians to diagnose and treat health conditions(Eyal, 2013).

In the client-based professions, the expert-client relationship is the locus of practice– and a sociologically rich nexus. A core feature of professional work is that, because the clientele that engages an expert generally does not comprise practicing members of the professional community, clients are non-expert in the services they require (Freidson, 1988). Therefore, clients confront the challenge of choosing providers and determining whether to follow expert advice even though they rarely possess the understandings to assess the quality of the guidance proffered.

The professions thus represent instances of decision making under asymmetric information, because the expert has far deeper knowledge than the client. In fact, although phrased differently, many scholars of the professions have regarded restricted access to esoteric knowledge as the most general, quintessential characteristic of a profession: to become an expert in a field is to apprentice in and master a complex, opaque and specific body of knowledge (Abbott, 2014). Professional work is grounded in this complex and exclusive system of specialized knowledge, which includes both a codified, abstract knowledge base that is acquired in formal educational settings, as well as experiential knowledge that is tacit in nature and developed through apprenticeship and on-the-job (Abbott, 2014; Freidson, 1988).

The asymmetry in knowledge between clients and experts necessitates trust in client interactions in the markets for expertise. As Hughes (1963, pp.656-657) described it, "Since the professional does profess, he asks that he be trusted. The client is not a true judge of the value of the service he receives; furthermore, the problems and affairs of men are such that the best of professional advice and action will not always solve them. ... . The client must trust [the expert's] judgment and skill." In the context we study, patient-physician dyads, Greenfield et al. (2012, p. 1203) further elaborate, "This dyad carries inherent elements of trust, loyalty, intimacy and dependency, that are rooted in the patient-physician emotional contract, and implies a strong interpersonal relationship. Trust and satisfaction are major predictors of patient loyalty and mutual commitment to treatment success."

Restating Hughes: a client generally is neither in a position to *ex ante* evaluate an expert's expertise, nor even to infer the quality of advice from *ex post* outcomes. Extending this at least a step further, work in symbolic interactionism questions whether expertise itself can ever be truly objective or fully codified? Rather, because expertise is situated and shared in embedded contexts and it is characterized by myriad evaluative uncertainties, the nuances in social interaction are part and parcel of it. This contrasts to a view of expertise as an objective resource that is created and controlled by the professions and dispensed in a more regularized, universal manner (e.g., Hughes, 1994; Barley, 1989; Becker, 1970). Moreover, it implies that not only is there a knowledge asymmetry between client and expert, but the alleged ground truths in a profession are just that: alleged. Of importance to our argument, this insight also connects the study of expertise to the enormous bodies of scholarship in the sociology of work, occupations, and economic markets.

For instance, a large, relevant literature documents how consumers respond to market uncertainty. In product and cultural markets, consumer evaluations of goods of uncertain quality often default to assessments of the identity and characteristics of producers, because the latter is far easier to assess than the former. Likewise in the markets for expertise. How do we know the quality of a thoracic surgeon when we, the patient, know nothing of the subject? In such situations, we evaluate more-concrete, easier-to-observe signals of the quality of a provider, such as the names on the framed degrees adorning the office walls during consultations, the stature of the hospital or practice or law firm that employs an expert, the referrals we receive from others whom we assume to be better informed than we are. And oftentimes, the literature shows, we also make assessments based on the gender of the expert.

More broadly, certain ascriptive and status characteristics, including gender, often influence how people assess ability even when there is no actual relationship between them and merit (Berger et al., 1972; Ridgeway and Erickson, 2000). This is known to be true in many facets of labor markets, even at the turnstiles of entry to employment in organizations, as non-merit-based criteria, including race and gender, influence whether and how job applicants are evaluated (e.g. Goldin and Rouse, 2000; Petersen et al., 2000; Castilla, 2008). In expert-client dyads, therefore, an individual's gender may elicit presumptions about expertise because this is a cultural context in which there are gendered assumptions about ability, regardless of their inaccuracy.

Within expert markets, there is evidence of gendered perceptions of ability. In one, telling study, Prince et al. (2006) surveyed patients in a hospital Emergency Department to ask whether they had been seen by a physician, when *all* patients had received a consultation. In patient-physician contacts, 93.3 percent of consultations with male doctors were recognized by patients as a physician visit, compared to only 78.5 percent of consultations if the physician was female. Thus, females were significantly less likely than males to be recognized as doctors, indicating the public's gendered perception of the role. In a study of online posts among economists, Wu (2020) found that discussions about women often described personal characteristics rather than their professional accomplishments. In general, posts on the forum about women were significantly more likely to fork conversation toward non-professional topics, relative to posts about men.

Among physicians themselves, there is significant evidence of gendered preferences in selection of providers. Studying referral networks in US healthcare, Zeltzer (2020) demonstrates significant gender homophily among physicians treating Medicare patients: providers are more likely to refer their patients to same-gender clinicians, even after careful adjustments for specialty, patient health conditions, and other potentially confounding factors. This article also shows that a large fraction of the earnings disparity between male and female physicians may be attributed to gender homophily in the physician referral network. Likewise, Sarsons (2017) finds that referring doctors interpret patient outcomes differently, depending on the gender of the specialist provider. In particular, referrers become more pessimistic about a female surgeon's ability than a male's following a patient's death, indicated by a sharper decline in future referrals to the female surgeon after an adverse event.

Sarson's findings in medicine illustrate a broader phenomenon demonstrated across occupational contexts, which is that ability is even more likely to be questioned if a worker is female when that person is a gender-atypical occupant of a professional role (Kanter, 2008; Ibarra, 1992; Ridgeway and Smith-Lovin, 1999; Ding et al., 2013). But what of the gender of the client or patient? Are assessments of the qualifications of a female provider likely to depend on the gender of the client? Using an audit study methodology, Greene et al. (2018) show that the answer to this question appears to be, yes. They recruited survey participants to select a physician based on randomized names, while fixing information about doctor quality. Survey respondents favored white and male names compared to African American male, African American female, or Middle Eastern names, but this was particularly so for *White and male* study participants. Likewise, Hall et al. (1994) is revealing. They found that patients who were examined by young physicians, especially if female, reported lower satisfaction ratings for the consultation. This finding was true for male and female patients alike, but the lowest absolute level of satisfaction was reported in male patient-female physician dyads.

These arguments lead us to three propositions about gender dynamics in markets for expert opinions. As we will elaborate below, we study the incidence rates of *second* opinions, because we believe these to be one indicator of the questioning of expertise.

*Proposition 1: In markets for expert services, clients are more likely to question the guidance of female experts than comparable male experts.*

*Proposition 2: Compared to female clients, male clients are more likely to question the guidance of female experts than comparable male experts.*

*Proposition 3: If a second expert opinion is sought in mixed-gender client-expert dyads, the client will be prone to consult a same-gender-as-the-client expert for the second opinion.*

# Context

We believe that antecedents to questioning expertise, especially those that rise to the level of a cause of clients' decisions to solicit a second opinion, are under-studied mechanisms for the gender pay gap in the high skill professions. In expert markets, a second opinion arises when clients seek the council of a second expert concerning the diagnosis of a situation, recommended course of action, or prognosis. We believe that second opinions in expert engagements are more likely to occur when the client lacks full confidence in the opinion of an expert–when the "trust" that lubricates exchange in these markets is not fully in tact. Insofar as there are predictable and consistent, gendered patterns in the incidence of client-initiated SOs, this phenomenon is likely implicated in gender-based disparities in occupational attainment among experts.

In healthcare specifically, second opinions (SOs) occur for a number of reasons: (i.) the nature of the medical condition, (ii.) physician preference, and (iii.) patient preference. For instance, in academic medical centers, second reviews of anatomic pathologic diagnoses are routine (Swapp et al, 2012, 8th citation in Payne). In complex clinical cases, specialists themselves may prefer that the patient seeks a second opinion, when a diagnosis or treatment is ambiguous or when the patient's condition is on the edge of a clinician's experience. However, in each of the first two motivations for a second opinion, a patient's confidence in physician expertise plays little role in the decision to solicit the SO. In the first instance, patients do not personally encounter pathologists or radiologists, so expertise is anonymously rendered; in the second setting, SOs are specialist-initiated. As described in detail below, we therefore exclude these cases from the analysis. Move the following: Depending on the study, the literature suggests that in 10% to 62% of cases, SOs lead to a significant change in diagnosis, treatment or prognosis (Payne et al. 2014).

A third category of second opinions are more likely to be *patient/client-initiated*, and these are the focus of our analysis. The existing literature on patient-initiated SOs, the vast majority of which is survey-based, indicates that patients pursue SOs for a few reasons, including that symptoms persist despite having sought medical care, that complications arise during treatment, and when the patient is dissatisfied with or lacks confidence in an expert consultation (Payne et al. 2014). Assuming a high incidence of the latter rationale, SOs reveal preferences held by patients and implicitly, the level of trust that patients place in the recommendations of their physicians.

Indeed, this interpretation of patient-initiated SOs is consistent with the small literature on physician reactions to them. Greenfield et al. (2012) found that physicians sometimes report feelings of resentment, disappointment, and embarrassment when they discover that a patient has sought a second opinion. Physicians also report that patients often attempt to conceal SOs from consulting physicians, creating two sets of patient-specialist pairs that rarely develop into a direct line of communication between the two, consulting physicians.

Below, we describe a strategy to identify patient-initiated second opinions from medical claims data and detail several analyses to validate these cases as SOs. We then document gender differences in SO rates in a very large sample of speciality visits.

# Data

We examine the three propositions in the Massachusetts All Payers Claims Database (MA APCD). The MA APCD is collected and maintained by the Center for Health Information and Analysis (CHIA) and contains remarkably comprehensive information derived from the medical and pharmacy claims of virtually every resident in Massachusetts. We possess these data for January 1, 2010, to December 31, 2015.

Massachusetts requires all health insurers in the state to report detailed information on every medical and pharmacy claim they receive. CHIA collects these data and prepares them for use in research. For instance, CHIA processes the data to create a hashed identifier to link records of individuals that change insurance plans over time. The MA APCD contains multiple data files, but we mainly draw from the medical claims file, which contains in the vicinity of 1 billion distinct medical claims. Data in the medical claims file include a physician identifier, a patient identifier, diagnosis codes, dates and locations of provider visits, identification of medical procedures performed, charged dollar amounts, and, importantly, referral information. The latter includes an indicator for whether the patient was referred to a given specialist and an identifier for the specific, referring physician.

# Sampling

Data as comprehensive as the the state All Payers Claims Database have only become available recently, and there is no widely agreed-upon method for identifying patient-initiated second opinions in large medical claims data sets. Shmueli et al. (2019), in the first article to propose a method for identifying second opinions in medical claims data, recently wrote, "To the best of our knowledge, all studies that evaluated SO utilisation so far were based on patient self-reported surveys and not on objective data, which makes it difficult to compare among different studies and countries". Even the survey-based literature on second opinions is small. In a comprehensive literature review, Payne et al. (2014) identified only 13 articles that met their inclusion criteria.

We define the risk set for second opinions to be the set of all first-time office visits to a specialist for a new health condition, for which the patient has no observed medical history. Each observation in this set represents an index visit and has the potential to result in a second opinion.

This definition is similar to Shmueli et al. (2019), which defined a "second medical opinion" to occur in all cases when a patient consults a second specialist, *in the same specialty as a first specialist*, within three months of the first consultation. Because the APCD data available to us appear to be substantially more detailed than those used in Shmueli et al. (2019), we have the opportunity to create a more precise sample of SOs.

To construct the sample of index visits, we leverage the referral indicator in the medical claims data and sample all instances in which an adult patient is referred by their PCP to a specialist in specialty $i$ for the first time.[1] We limit the index sample to *office visits* to specialists (CPT codes 992**) because our goal is to identify cases in which a patient consults a specialist to obtain a diagnosis and treatment recommendation. This sampling strategy explicitly excludes several scenarios. First, by sampling only the first time a patient sees a physician in specialty $i$ we effectively limit our index visits to specialty-naive patients. This ensure that the index visit is not itself an SO, and it reduces variation in patients' specialty-specific knowledge.

Second, by restricting the index sample to office visits, we eliminate specialties such as radiography and pathology in which patients and experts rarely meet in person. Third, by sampling office visits resulting from a referral from the PCP, we remove cases that begin with an emergency department visit where there is limited scope for patients to choose which physician they consult. Finally, by removing patients younger than 18 years old, we exclude cases where the patient is not the primary decision-maker.

The resulting sample includes almost 1.6 million referrals from PCPs to specialists that resulted in a first-time office visit of the patient to a specialist in field $k$. Descriptive statistics for the sample, broken out by the gender of the specialist in the index visit, are shown in table 1.

In table 1, several differences between male and female specialists stand out. The first few rows of the table show substantial gender sorting in the selection of specialists. Both female patients (row 2) and female PCPs (row 4) choose female specialists more often. Since female specialists have, on average, entered the workforce more recently than male specialists (row 3) and because there is age homophily between patients and specialists, female specialists see slightly younger patients (row 1). The Charlson score, which is a standard index of comorbidities that summarizes overall patient health, is higher for male specialists. Therefore, on average, male specialists see more seriously ill patients, which is mostly explained by gender sorting among specialties. In the detailed breakdown of specialty by gender, we see that male and female physicians specialize in substantially different fields. For example, women are over-represented in dermatology, while men are over-represented in orthopaedic surgery. Finally, Table 1 also shows that patients seen by female and male specialists are relatively similar in terms of the insurance plans that cover their patients.

Next we indicate whether an index visit resulted in an SO. We do so by identifying index visits in which the patient is *referred by their PCP* to a second specialist in the same speciality within 180 days of the index visit.[2] To be precise, we define a second opinion to

[1] The referral indicator is used in all HMO and POS insurance plans. It also is populated in a few PPO plans in Massachusetts. The indicator identifies the physician that provided the referral.

[2] We selected 180 days because it often takes months for a new patient to obtain a first-visit appointment with a specialist. In sensitivity analyses, we re-estimate our models using different time windows and find highly comparable results.

Table 1: Descriptive statistics by gender of the specialist in the index visit

| | Male specialist | Female Specialist |
|---|---|---|
| Patient age (Mean) | 48.94 | 46.51 |
| Female patient (%) | 53.93 | 72.00 |
| Specialist graduation year (mean) | 1986 | 1993 |
| Female PCP (%) | 39.47 | 53.66 |
| Charlson index score (Mean) | 0.62 | 0.53 |
| Insurance type (%) | | |
| Health Maintainance Organization (HMO) | 72.19 | 72.75 |
| Point of Service (POS) | 7.79 | 8.63 |
| Preferred Provider Organization (PPO) | 6.99 | 7.95 |
| Medicaid | 3.04 | 2.66 |
| Exclusive Provider Organization (EPO) | 1.87 | 1.81 |
| Other insurance type | 8.12 | 6.19 |
| Provider specialty (%) | | |
| Dermatology | 16.05 | 34.56 |
| Orthopaedic Surgery | 16.20 | 2.96 |
| Otolaryngology | 7.51 | 5.25 |
| Urology | 8.10 | 1.68 |
| Surgery | 6.53 | 7.00 |
| Gastroenterology | 7.58 | 4.79 |
| Ophthalmology | 6.07 | 5.22 |
| Neurology | 5.07 | 5.84 |
| Cardiovascular Disease | 6.15 | 2.73 |
| Obstetrics and Gynecology | 1.67 | 9.47 |
| Other specialty | 19.07 | 20.50 |
| Observations | 1,210,491 | 385,778 |

Notes: This sample includes only index visits by adult patients who were referred by their PCP and had not previously seen a specialist in the focal specialty.

occur when, conditional on having not consulted a specialist in field $k$ in the past, a patient consults two physicians in the same medical specialty $k$ in a 180-day window, and both appointments were established by a referral from the patient's primary care physician. We describe these data in table 2. The table shows that female patients and female specialists are overrepresented in index visits that result in SOs compared to the majority, non-SO cases. Table 2 also shows that some medical specialties, such as orthopaedic surgery, are over-represented in second opinion cases while others are under-represented. Finally, the table shows that about 4% of all visits result in a second opinion.

## Validating the second opinion label

Because there is limited literature and no external validation of second opinions (versus first-time visits for a second medical condition) as we have defined them here, we perform several analyses to establish that the SO we have identified have characteristics that systematically

Table 2: Descriptive statistics by whether the index visit results in a second opinion

| | Second Opinion | No Second Opinion |
|---|---|---|
| Patient age (Mean) | 47.77 | 48.37 |
| Female patient (%) | 59.31 | 58.25 |
| Specialist graduation year (mean) | 1988 | 1987 |
| Female specialist (%) | 25.55 | 24.11 |
| Female PCP (%) | 43.92 | 42.86 |
| Charlson index score (Mean) | 0.67 | 0.59 |
| Insurance type (%) | | |
| Health Maintainance Organization (HMO) | 73.13 | 72.29 |
| Point of Service (POS) | 8.60 | 7.97 |
| Preferred Provider Organization (PPO) | 6.83 | 7.24 |
| Medicaid | 2.84 | 2.95 |
| Exclusive Provider Organization (EPO) | 1.48 | 1.87 |
| Other insurance type | 7.11 | 7.68 |
| Provider specialty (%) | | |
| Dermatology | 19.91 | 20.55 |
| Orthopaedic Surgery | 17.87 | 12.81 |
| Otolaryngology | 6.75 | 6.97 |
| Urology | 7.94 | 6.49 |
| Surgery | 5.59 | 6.68 |
| Gastroenterology | 4.46 | 7.00 |
| Ophthalmology | 9.16 | 5.73 |
| Neurology | 5.14 | 5.26 |
| Cardiovascular Disease | 5.83 | 5.30 |
| Obstetrics and Gynecology | 5.56 | 3.48 |
| Other specialty | 11.78 | 19.72 |
| Observations | 60,557 | 1,535,712 |

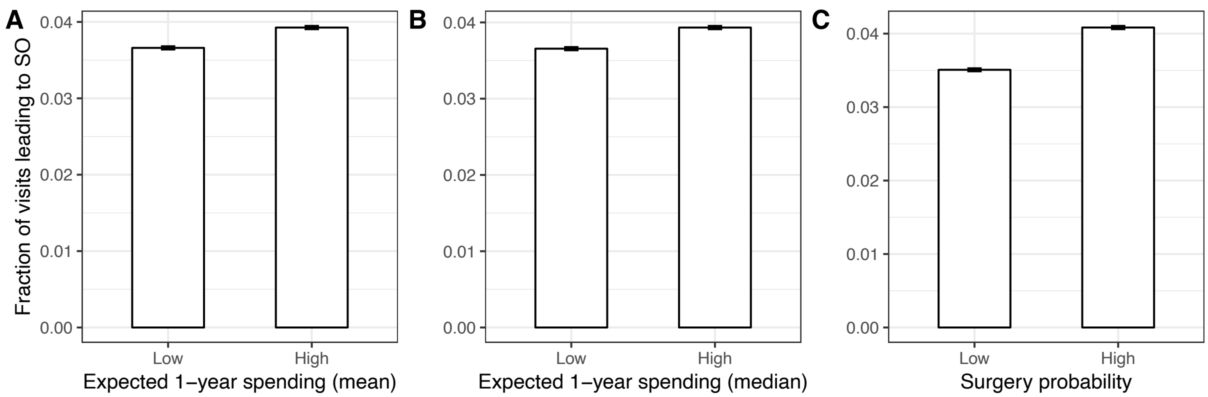
Notes: This sample includes only index visits by adult patients who were referred by their PCP and had not previously seen a specialist in the focal specialty.

distinguish them from other specialist visits.

First, based on the existing, survey-based data examined in prior work, we expect that SOs should be more prevalent for serious health conditions. To assess this, we compute the average for two proxies of severity of a health condition – the 1-year medical spending and 1-year surgery risk associated with every *diagnosis*. In other words, this analysis is performed at the level of the health condition. The intuition for this exercise is that patient-initiated SOs should be more common if the medical condition diagnosed in the first visit is costly or complex to treat or has a high probability of requiring surgical intervention.

To compute expenditures, we identified all instances in which a Massachusetts resident $j$ was first assigned diagnosis $i$. We then aggregated all allowed medical expenses in the following year and computed the mean and the median of this sum, for all diagnoses. Likewise, we compute a 1-year surgery probability using the same strategy but based on whether

Figure 1: Second opinions and the severity of a diagnosis.




Note: This graph shows the fraction of index visits leading to an SO, conditional on the expected 1-year spending an surgery probability associated with the main diagnosis given at the first visit. Expected 1-year spending is computed by taking all first instances in which a patient receives a diagnosis across the MA APCD (so not just limited to our index visits) and then aggregating medical spending of that patient in the following year. In panel A, we assign an index visit the mean 1-year spending for a diagnosis, in panel B we use the median spending. Surgery probability is computed in the same way – it represents the probability that patient $i$ diagnosed with diagnosis $j$ will undergo surgery in the year following the diagnosis. For both expected spending and surgery probability, we take the median and split the sample into "high" and "low" cases. An index visit with high expected spending means that the diagnosis given in that index visit is associated with above median expected spending. An index visit with high surgery probability means that the diagnosis given in that index visit is associated with above median likelihood of surgery in the following year. Each bar includes an error bar representing the standard error.

patients actually had surgery in the year following the initial diagnosis.[3] The results are shown in figure 1. As anticipated, when expected spending and surgery probability are high, the probability that the patient will received a second opinion is substantially higher.

An unknown fraction of SOs occur because there is a mismatch between the expertise of the first-seen physician and a patient's health condition. There can be elaborate sub-specializations within the major medical fields, and patients occasionally may be mis-assigned to specialists in the referral process. Note that in constructing the set of index visits, this is why we exclude all SOs that originate from a referral that was made by a specialist provider, rather than a PCP. In a small number of cases, however, it is possible that a patient's PCP is the referrer of record for insurance purposes, even if the suggestion to see a different physician originated from the specialist in the first visit.

To evaluate the possibility that the sample of "patient-initiated" SOs arises because of a large fraction of expertise mismatches to patient conditions, we compare the exact treatment histories of the specialist in the first opinion visit to that of the specialist in the SO consultation. The intuition for this analysis is that, if patients frequently are referred to specialists that do not have the expertise to treat their conditions, the experience distribution of the first- and second-opinion specialists should be more dissimilar than pairs of same-specialty physicians that we will choose at random. Conversely, if the pair of physicians handle similar

[3]We used the 'narrow' definition developed by HCUP to identify CPT codes that indicate surgical interventions: https://www.hcup-us.ahrq.gov/toolssoftware/surgeryflags_svcproc/surgeryflagssvc_proc.jsp.

cases (and therefore represent suitable alternatives for first and second opinions for a given diagnosis), we should observe that the exact treatment histories of the pair of specialists are considerably more similar than are those of two, randomly selected providers.

We conduct this analysis by extracting in the SO sample the complete, 1-year treatment history of the first opinion (FO) and second opinion specialist. This yields a vector of frequencies of procedure codes performed by each specialists. We then compute the cosine similarity between the experience vectors of the FO and SO specialists. To benchmark the resulting distribution of similarities we also compute distances between the 1-year treatment history of the FO specialist and two, alternatively sampled specialists. In one approach, we randomly sample a specialist in the same speciality who also treated patients on the same day as the index visit. In a second approach, we match specialists on the diagnosis that they most commonly treat and then randomly sample a specialist in the same speciality who also treated patients on the same day as the index visit.

The three resulting distributions of similarity scores are shown in figure 2. Here, the results are truly remarkable. We find that on average, the actual pair of FO and SO specialists is much more similar than the pairs with counterfactually assigned specialists placed into role of the SO provider. This persuasively demonstrates that on average, patients seek SOs from specialists with expertise in the same sub-specializations as the physician in the index / FO consultation. Likewise, the presence of highly overlapping expertise in the realized FO-SO pairings shows us that mismatches between specialist expertise and health condition in the index visit is unlikely to be common, and patient mis-assignment is not a central part of the data generating process in the sample of SOs.[4]

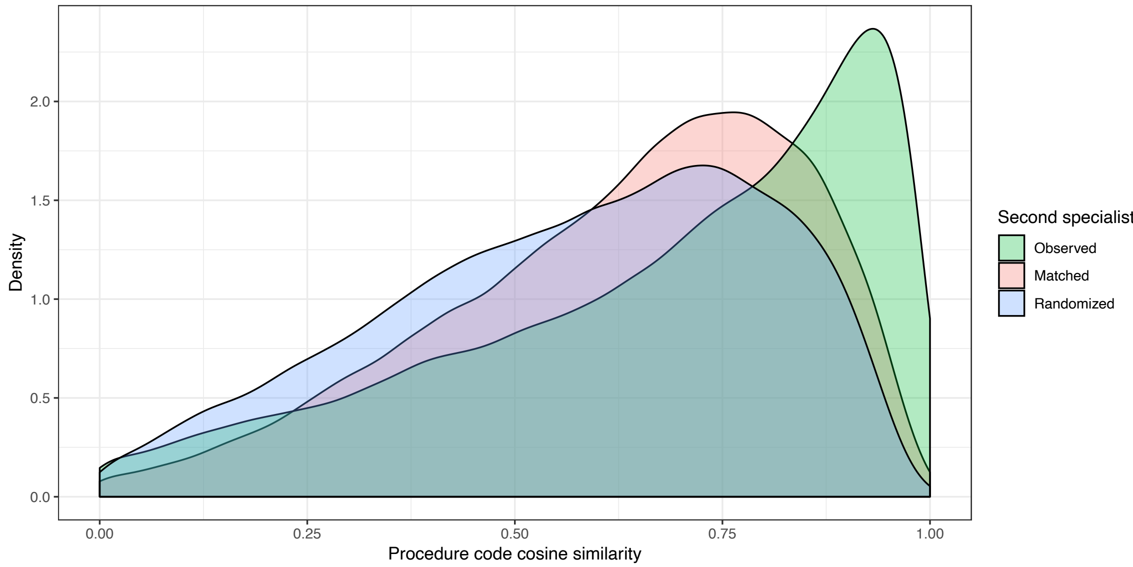
The third analysis we conducted to validate the sample of SOs considers the possibility that a new health issue emerged between the first and second specialist visits. If this occurs frequently, it will conflate the sample of SOs. A patient that experiences a second health problem in the specialty of the first provider and then chooses a second clinician is not seeking a SO for the original medical consultation. We therefore examine the rate at which a patient in our SO sample, (i.) visits their PCP between consultations 1 and 2, and (ii.) is assigned a new diagnostic code that differs from the first specialist visit.[5] We find a PCP-initiated diagnosis code change in 7% of the cases in the SO sample. Note that this includes many cases when the PCP simply inputs a diagnosis code for the same health issue but is less specific than the one assigned by the specialist. Given the low percentage of cases in which there is a code update, this analysis bolsters confidence in the accuracy of our procedure for identifying in SOs.

In the analyses to follow, we use the sample of patient-initiated SOs described in table 2. However, we conduct multiple sensitivity checks that exclude the cases that, based on our previous two analyses, are possible, false positives. Specifically, we exclude cases we label as SOs but could be a new health concern, and we also exclude cases in which the randomly

[4]In the Robustness Checks section of the paper we describe an additional, empirical strategy to evaluate the impact of potential mismatches between patient health conditions and specialist expertise in the index visit. That analysis also will show that the central empirical results in the paper are highly robust to this concern.

[5]We restrict these diagnoses to those that are reasonably likely to relate to a health issue that can be treated in the focal speciality. Specifically, the diagnosis needs to occur at least once in every 1,000 visits to a physician in that specialty.

Figure 2: Distribution of cosine similarities between observed, matched and randomly assigned pairs of specialists.




Note: This graph shows the distribution of similarities between the treatment histories of the first and second specialist in green. The two counterfactual scenarios are shown in red and blue. Compared to the counterfactual distributions, the observed distribution is shifted to the right and clearly peaks close to 1, suggesting higher levels of similarity in actually occurring physician pairs.

imputed pair of specialists is more similar than the actual pair in the first-second opinion specialist dyad. Finally, we reduce the time window until a SO appointment from 180 to 90 to 30 days.

# Evaluating baseline homophily

(Need to motivate this better) We use the full sample of 1.6 million first specialist visits described in table 1 to estimate a dyadic choice model to establish the level of gender homophily in first visits of patients to specialists. This analysis reveals the baseline tendency for gender homophily in the Massachusetts medical market, and we include it here because it is of interest in its own right, but more directly, it guides our interpretation of subsequent results.

Both patient and PCP are likely to influence the choice for a specific specialist, so we account for gender pairings between both the PCP and specialist and the patient and specialist. We estimate gendered preferences of both patients and PCPs using a conditional logit model for the probability that PCP $j$ or patient $i$ selects focal specialist $k$. We limit the identifying variation to differences within each patient's and PCP's choice set. As a result, any patient or PCP-level attributes are conditioned out of the analysis.

The data consist of an observation for each dyad (ij, k), with specialist and dyad (pairwise) characteristics $X_{ik}$, $X_{jk}$, and a binary outcome that indicates whether patient $i$ visited specialist $k$. In other words, each referral from a PCP to a specialist generates a realized

PCP-specialist and patient-specialist dyad. For each, realized referral, we create a set of counterfactual matches that consists of specialists that the PCP could have referred to and the patient could have visited. To establish the risk set of counterfactual matches, we compare each chosen specialist $k$ with not-chosen specialists $k$ from the same Hospital Referral Region (HRR), from the same specialty of $k$ and that practiced in the same month as the specialist to which patient $i$ was referred by PCP $j$. This counterfactual choice set makes a weak assumption about substitutability. Specifically, specialists in the same city and medical specialty are not assumed to be perfect substitutes. Rather, we assume only that specialists from different markets or from different medical specialties are not in the choice set. Variation in covariates that are included as controls (e.g., geographic distance between patient and specialist) capture gradations in substitutability.

The results are shown in table 3. Estimating the conditional logit is computationally intensive because of the number of potential, non-occurring dyads. To keep the number of observations in each regression below 100 million, we estimate the matching model by year rather than in pooled cross sections.[6] The results are very similar across years. The table suggests that, all else equal, patients are about 20% more likely to visit a specialist of the same gender. PCPs exhibit weaker gender preferences but are still 12 to 15% more likely to refer to a specialist of the same gender, conditional on the gender of the patient. When jointly interpreting the Same gender (Patient) coefficients and the Female specialist coefficients, the models suggest that while female patients are slightly more likely to select a male specialist over a female specialist, *the difference is substantially more pronounced for male patients*. Finally, physician experience and distance also are important determinants of referrals. The experience gradient suggests an inverted U shape in which mid-career specialists are in highest demand. All specifications of distance reveal a sharp decline in match probabilities as a function of geographic separation.

In sum, the baseline estimates show a strong homophilous preference in physician choice, especially among male patients. This result fully conditions on the gender distribution of available, nearby physicians in each specialty. This finding is important in interpreting the significance of the results that follow, because gender sorting in the first-stage matching process for first opinions stack the deck against finding any gender differences in SOs. This is because male patients that hold a strong preference for a male specialist have selected a male provider *in the first stage of the referral process*. When we observe male patient-female specialist dyads in first visits, it is safe to assume that the majority of male patients in these dyads have a weaker-than-average-male preference for a same-gender / male provider. The bottom line is that the analysis that follows likely provides a conservative estimate of the effect of patient-specialist gender matches on patient-initiated SOs.

# Research design

In our main specifications, we estimate the dependence of the rate of second opinions on the genders of the specialist and patient. A primary concerns in such a specification is that one or more unobserved confounder may be correlated with specialist or patient gender and the

[6]The sample size declines each year because we restrict the data to include only patients that visit a specialist in a specialty for the first time within the data. We return to this choice in the Robustness section.

Table 3: Baseline homophily first visit

| Dependent Variable: | Patient visited specialist/PCP referred to specialist | | | | |
|---|---|---|---|---|---|
| | 2010 | 2011 | 2012 | 2013 | 2014 |
| | (1) | (2) | (3) | (4) | (5) |
| *Variables* | | | | | |
| Female specialist | -0.2813*** | -0.2674*** | -0.2357*** | -0.2555*** | -0.2384*** |
| | (0.0043) | (0.0053) | (0.0056) | (0.0058) | (0.0065) |
| Same gender (Patient) | 0.2355*** | 0.2318*** | 0.1966*** | 0.2019*** | 0.1894*** |
| | (0.0045) | (0.0055) | (0.0058) | (0.0061) | (0.0068) |
| Same gender (PCP) | 0.1529*** | 0.1228*** | 0.1291*** | 0.1255*** | 0.1437*** |
| | (0.0042) | (0.0052) | (0.0055) | (0.0058) | (0.0065) |
| Grad. year pre 80 | 0.2395*** | 0.0671*** | -0.0028 | -0.0538*** | -0.1031*** |
| | (0.0056) | (0.0062) | (0.0065) | (0.0067) | (0.0075) |
| Grad. year 80-00 | 0.4113*** | 0.2453*** | 0.1933*** | 0.1488*** | 0.1065*** |
| | (0.0051) | (0.0055) | (0.0056) | (0.0056) | (0.0060) |
| Distance (Km) | -0.0866*** | -0.0862*** | -0.0829*** | -0.0819*** | -0.0821*** |
| | (0.0003) | (0.0004) | (0.0004) | (0.0004) | (0.0004) |
| *Fit statistics* | | | | | |
| # of visits | 468,259 | 312,457 | 268,130 | 241,147 | 190,830 |
| Observations | 94,729,369 | 60,668,246 | 49,669,119 | 43,812,135 | 33,301,615 |
| Pseudo $R^2$ | 0.15167 | 0.14842 | 0.14696 | 0.14613 | 0.14664 |

Notes: All models are estimated using conditional logits.
$^{*}$p<0.05; $^{**}$p<0.01; $^{***}$p<0.001 (Two-tailed tests).

error term (i.e. $corr(X_i, u_i) \neq 0$). The most obvious candidates for an omitted variable is the unobserved quality of the specialists in the data, perhaps followed by the health status of patients. While we are able to include a rich set of controls based on the track record of the specialist and the health history of each patient, conditioning on these variables may not fully satisfy the assumption that $corr(X_i, u_i) = 0$.

Our best, available strategy to address this concern is to estimate models using physician fixed effects (FEs). To implement the FE estimator, we run two separate regressions, one for female and one for male specialists. These specifications allow us to ascertain whether male and female patients request SOs from the *same specialist* at different rates. In other words, all within-year characteristics of the physician, including quality, are held constant in these regressions. Because most patients are observed only once, we cannot include patient fixed effects. In addition to the FE specification, we also leverage the idea that if SO rates were explained by differences in health status between male and female patients, then, all else being equal, the decisions to request a SO should be independent of the gender of the specialist seen first.

# Results

Table 4 reports six Ordinary Least Squares (OLS) regressions in which the outcome is a second opinion. The regressions include: i.) a full vector patient age dummies, ii.) a vector of specialist graduation year dummies to capture first opinion clinicians' experience, iii.) the Charlson score, based on the 1-year medical history of the patient, and iv.) a complete set of insurance plan indicator to fully capture all differences in patient copayments and reimbursements.

Col 1 includes only the gender of the specialist and shows that patients of female first opinion specialists request a SO more frequently than those of male specialists. The direction of the effect remains the same after including specialty and year FEs, but the magnitude of the coefficient increases markedly after including specialty fixed effects. This occurs because the highest incidence of SOs are among (e.g.) candidates for orthopaedic surgery, and specialized surgeons are much more likely to be male. Models 4 and 5 include the gender of the patient and an interaction between gender of the patient and gender of the first opinion specialist. The results show that female patients request second opinions more frequently than male patients but that the decision-making of male patients is much more sensitive to the gender of the specialist. In model 6, we include PCP fixed effects, which do not alter the results.

Table 4: Second opinion rate and gender

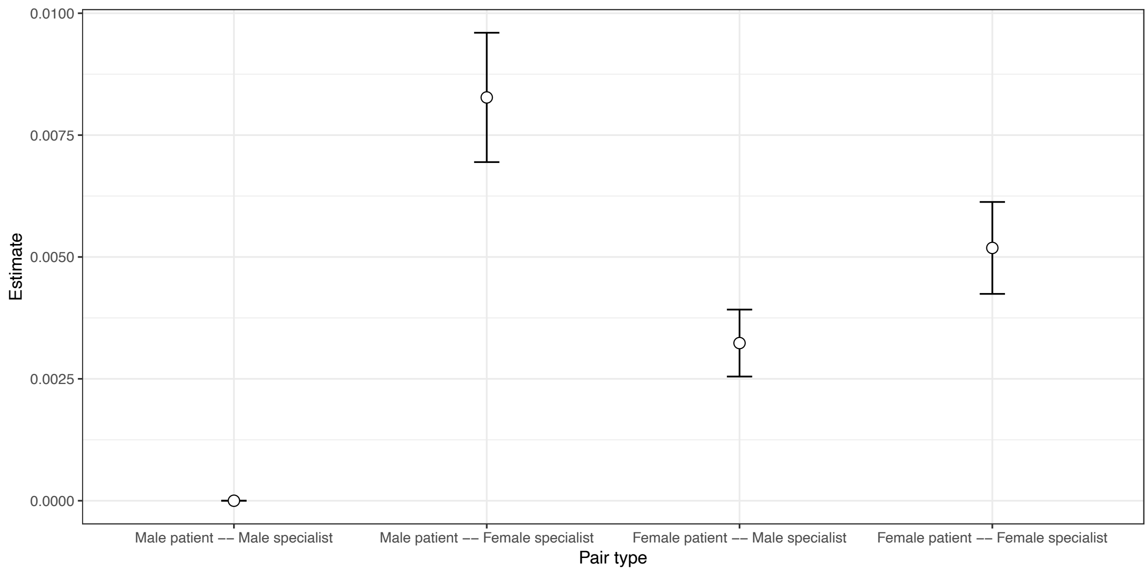
| | *Dependent variable:* | | | | | |
|---|---|---|---|---|---|---|
| | Probability of second opinion after first consult | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Female specialist | 0.002*** | 0.004*** | 0.004*** | 0.004*** | 0.008*** | 0.008*** |
| | (0.0004) | (0.0004) | (0.0004) | (0.0004) | (0.001) | (0.001) |
| Female patient | | | | 0.002*** | 0.003*** | 0.003*** |
| | | | | (0.0003) | (0.0004) | (0.0004) |
| Female specialist*Female patient | | | | | −0.006*** | −0.006*** |
| | | | | | (0.001) | (0.001) |
| Constant | 0.043*** | 0.011* | 0.012* | 0.011* | 0.010 | |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) | |
| Specialty FEs | No | Yes | Yes | Yes | Yes | Yes |
| Year FEs | No | No | Yes | Yes | Yes | Yes |
| PCP FEs | No | No | No | No | No | Yes |
| Observations | 1,596,269 | 1,596,269 | 1,596,269 | 1,596,269 | 1,596,269 | 1,596,269 |

Notes: All models are estimated using OLS. Logistic regressions yield similar results.
*p<0.05; **p<0.01; ***p<0.001 (Two-tailed tests).

Figure 3 graphically illustrates the central pattern of results. It reports the point estimates for each of the four possible gender pairings based on the coefficients in Col 5, table 4. The graph shows that the SO rate for male patients varies much more than it does for female patients, depending on the gender of the specialist. The male patient - male physician combination is the reference group in the regressions and in the figure, and it represents the dyadic composition with the lowest rate of SOs. As we had anticipated, the greatest incidence of second opinions is in male patient-female physician dyads.

One, potential concern is that the patterns shown in table 4 may be driven by differences

Figure 3: Estimated effect sizes – extensive margin




Note: Figure shows the change in SO rate by gender-pair type: Female specialist - female patient, Female specialist - male patient, Male specialist - female patient and Male specialist - male patient. Male-male is the reference group. The graph shows that the behavior of male patients is much more affected by the gender of the specialist.

in specialist quality. To address this issue we estimate two regressions, one for male and one for female specialists, that includes specialist fixed effects. The results are shown in table 5. Although the coefficient estimates are similar, the effect sizes are different because the probability of a patient-initiated SO is substantially more likely when the specialist in the first visit is female. Specifically, if the specialist is female, female patients are 7.1% less likely than male patients to obtain an SO. If the specialist is male, male patients are 9.1% less likely than female patients to request a SO. Again, we find that the decision of male patients to pursue a SO is more strongly guided by the gender of the expert than is the decision of female patients, even when we limit variation to the gender of the patient while including a fixed effects for all specialists.

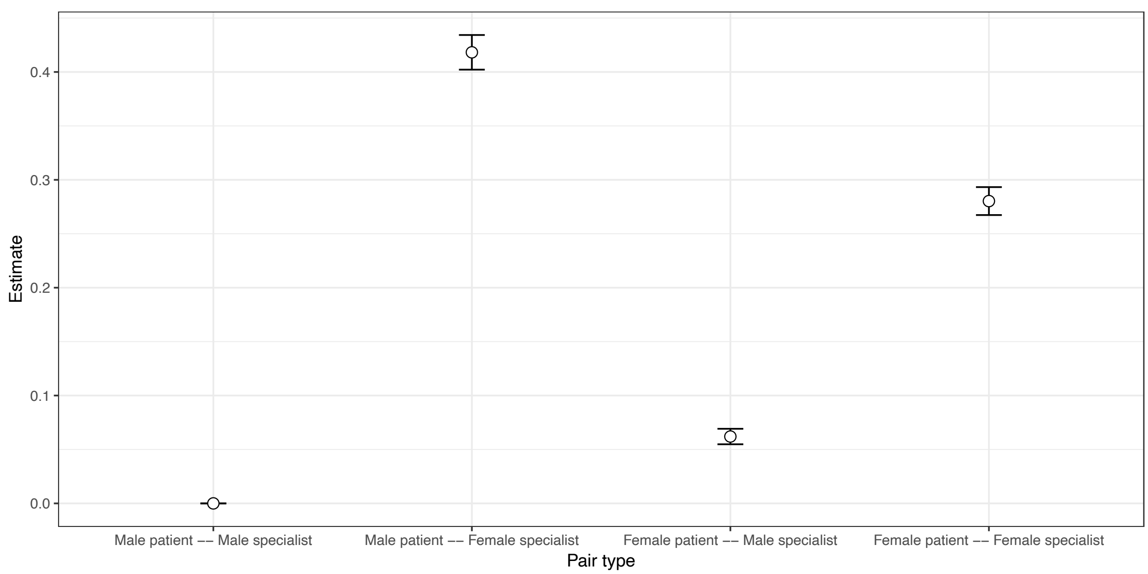
Next, we examine whether, conditional on seeing a specialist for an SO, patients alter their choice about the gender of the specialist. We are particularly interested in the following thought experiment: conditional on obtaining an SO after seeing a male or female specialist in a first visit, do male and female patients switch the gender of the second expert at different rates? Table 6 shows the regression coefficients of a model predicting the probability that patients obtaining an SO see two specialists with different genders across the two visits. The table shows that there is much less gender switching when the first expert is male compared to female. It also shows that men are much more likely to switch specialist gender for the SO than are women, *if the specialist in the first visit was female.* And although female patients also are more likely to switch gender when the first specialist is female, the discrepancy is much less pronounced compared to male patients. These differences are visualized in figure 4. In sum, this evidence suggests that patient decisions to request an SO are less driven

Table 5: Second opinion rate and gender

| | *Dependent variable:* | |
|---|---|---|
| | Probability of second opinion after first consult by specialist gender | |
| | Female specialist | Male specialist |
| | (1) | (2) |
| Female patient | $-0.003^{**}$ | |
| | (0.001) | |
| Male patient | | $-0.003^{***}$ |
| | | (0.0004) |
| Specialist FEs | Yes | Yes |
| Year FEs | Yes | Yes |
| Observations | 385,778 | 1,210,491 |

Notes: Both models are estimated using OLS. Logistic regressions yield similar results.
*p<0.05; **p<0.01; ***p<0.001 (Two-tailed tests).

Figure 4: Estimated effect sizes – intensive margin




Note: This graph shows gender switching for the sample of second opinions, for all gender pairings. It shows that male patients seeing female specialists for a first visit, are by a wide margin the most like;y to switch to a male clinician for an SO.

by gender when the specialist is male. It is also consistent with men holding much stronger preferences for male specialists than women hold for either male or female specialists.

Finally, we conduct a back-of-the-envelope calculation to assess the financial implications of the findings for specialists' earnings. When a patient obtains a SO, the patient may elect

Table 6: Gender switching in second opinions

| | *Dependent variable:* | |
|---|---|---|
| | Probability of switching specialist gender | |
| | (1) | (2) |
| Female specialist | 0.287*** | 0.418*** |
| | (0.005) | (0.008) |
| Female patient | | 0.062*** |
| | | (0.004) |
| Female specialist*Female patient | | −0.200*** |
| | | (0.010) |
| Constant | −0.006 | −0.052 |
| | (0.065) | (0.065) |
| Specialty FEs | Yes | Yes |
| Year FEs | Yes | Yes |
| Observations | 60,557 | 60,557 |

Notes: Regressions include a full set of specialist fixed effects. Column 1 is for female first opinion specialists. Column 2 is for male first opinion specialists. Both models are estimated using OLS.
*p<0.05; **p<0.01; ***p<0.001 (Two-tailed tests).

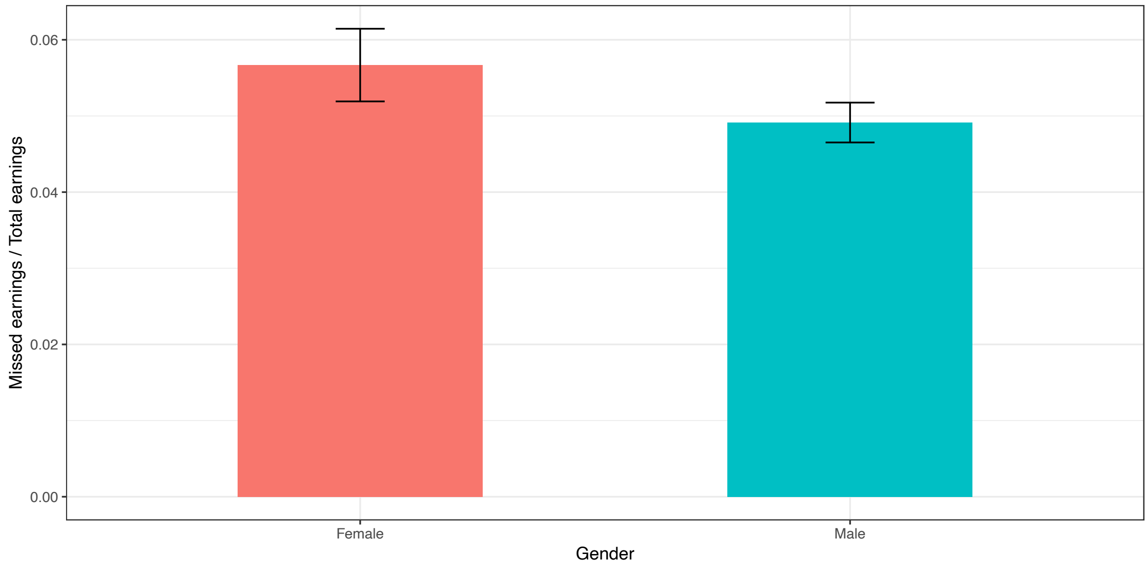
to return to the original specialist for treatment. However, the data indicate that this is uncommon, which is consistent with the view that SOs occur when patients lack full trust in the first opinion expert. On average, patient spending on services from the SO specialist is 80% higher than spending on services from the FO specialist, if the patient obtains an SO. Also, if we compare spending on services from specialists in a first visit between patients that later sought a SO and those that did not, specialists in non-SO cases receive 42% more in patient billings.

These findings indicate that spending on a second specialist are akin to "missed earnings" for the first specialist. If we assume that patient spending that shifted to the second specialist represents the missed earnings by the first specialist and if we express these missed earnings as a fraction of the total earnings of the first specialist (because women earn much less than men), we find that the missed earnings for female physicians as a fraction of their total earnings are 15% higher than the missed earnings for male physicians because of patient-initiated SOs. These differences are shown in figure 5.

# Robustness checks

We consider robustness checks and extensions, to assess the fragility of the findings and to shed additional light on interpretation. Beginning with the latter, we have argued that a significant fraction of patient-initiated SOs occur because the client lacks full confidence in

Figure 5: Differences in missed earnings




Note: This graph shows the difference in "missed earnings" expressed as a fraction of total earnings for female specialists compared to males. The 15% difference in the mean is statistically significant (t = 2.72).

the guidance of the first expert, based on the specialist's gender. An alternative possibility that we must consider is that certain patient-clinician gender pairing in fact result in lower quality medical care, and therefore obtaining SOs *should* depend on patient-physician gender pairings.

To explore this possibility, table 7 reports the difference in 30-day hospital readmission rates for both male and female patients, based on the gender of specialists. Readmission rates are a common measure of health outcomes and have been targeted by policy interventions in an attempt to reduce costs of health care (Ody et al., 2019). Here, we examine differences in adjusted 30-day readmission rates between male and female specialists. We record a readmission for every index visit by identifying patients who were admitted to a hospital in the year following the index visit and were readmitted to the hospital within 30 days of the initial admission date. We condition on patient age and the Charlson comorbidity score and we include provider specialty fixed effects.

In column 1, we compare readmission rates between male and female specialists. It shows that patients of female specialists are 9% *less* likely (the base rate is 1.5%) to experience a readmission in the year following the index visit. Female patients are 8% less likely than male patients to be readmitted in the year following the hospitalization. Model 2 includes an interaction effect between the gender of the specialist and the gender of the patient. The results show that female patients treated by female specialists (compared to female patients seen by male specialists) are less likely to experience hospital readmissions. Male patients, however, experience indistinguishable readmission rates when treated by a female specialist compared to treatment from a male specialist. In model 3 and 4, we examine the robustness of these results by (i.) limiting the sample to only index visits that did not lead to an SO, and (ii.) limiting the sample to only specialists that had at least one SO event in the sample.

The results are unchanged from model 2. In sum, the evidence leans toward opposition to the possibility that male patients' preference to match to male specialists leads to improved health outcomes.

Table 7: Readmission rates by specialist gender

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Probability of 30-day readmission in year following specialist visit | | | |
| | Full sample | | Non-SO visits | Specialists w/ SOs |
| | (1) | (2) | (3) | (4) |
| Female specialist | $-0.001^{***}$ | 0.0001 | $-0.00003$ | 0.0001 |
| | (0.0002) | (0.0004) | (0.0004) | (0.0004) |
| Female patient | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{**}$ |
| | (0.0002) | (0.0002) | (0.0002) | (0.0002) |
| Female specialist * Female patient | | $-0.002^{***}$ | $-0.002^{***}$ | $-0.002^{***}$ |
| | | (0.0005) | (0.0005) | (0.0005) |
| Observations | 1,596,269 | 1,596,269 | 1,535,712 | 1,533,424 |

Notes: All models are estimated using OLS. Logistic regressions yield similar results. Model 4 uses gender of the firsty specialist, model 5 uses gender of the second specialist.
$^{*}$p<0.05; $^{**}$p<0.01; $^{***}$p<0.001 (Two-tailed tests).

In table 8 we replicate results for the core regression analysis presented in table 4, model 5, but we exclude the observations that we believe to be at highest likelihood of being miss-classified as SOs. Specifically, model 1 removes all cases for which the actual pair of first- and second-opinion specialists are less similar in clinical expertise than is a randomized, imputed match to the first specialist, based on similarities in the actual, prior experience distributions among providers.

(NOTE: I THINK WE SHOULD DROP PHYSICIAN COMPARE) In model 2, we repeat this analysis but leverage differences between two specialty taxonomies. The specialty taxonomy used throughout the paper comes from the Centers for Medicare and Medicaid Services (CMS) and by our research design, the specialist in the first and second visit have the same specialty. Physician Compare, which is another data source that provides specialty data, gives physicians the opportunity to list more fine-grained specialty data. We merge the Physician Compare data with the Massachusetts APCD and we then exclude all cases where specialties are not the same for the first- and second-opinion specialist. The results, reported in model 2 of table 4, are similar.

In model 3, we exclude cases where patients (I.) visit their PCP between the first and second specialist consultation, and (ii.) are assigned a different diagnosis that is plausibly treated by a provider in the specialty of the first opinion. These cases may represent first opinions for a new condition, from a second specialist.

In models 4 and 5, we revisit the length of time until an SO. In (4), we define a SO as above but limit it to occur within three months after the first consultation. In (5), we further reduce the time to one month after the original consultation. While effect sizes change slightly, the central results are fully stable across the specifications in models 3-5.

Next, we explore potential problems associated with left censoring. The Massachusetts APCD data begin on January 1, 2010. For individuals that have their 'first' specialist visit in 2010, we do not have a full year of data to establish that they have not previously seen a specialist in the focal specialty and we lack the data to compile a one-year medical history to capture health status. We therefore re-estimated table 4, model 5 but we excluded all cases in which the first visit took place before January 1, 2011. Note that this removes a large number of cases from our sample. This is a byproduct of our research design because our sampling restrictions (i.e., the index visits contain only first visits to a specialty) front load many of our cases in the sample. However, despite the large reduction in sample size, the effect sizes remain remarkably stable.

Finally, to the extent possible, we consider the issue of external validity. Of course, the market for medical services differs from other markets for expert advice. Medical consultations are uniquely personal, and sometimes involve health concerns in which embarrassment and other emotional responses are commonplace. This raises the possibility that the pattern of results we observe is less indicative of questioning expertise, but instead has something to do with patients' comfort in ongoing consultation or treatment with a physician of a given gender.

To explore this possibility, we limit the index sample to first visits in which the diagnosis is approximately equally likely for male and female patients. In other words, we eliminate all single-sex-concentrated health problems. The table X regression limits the sample to first consultations in which the assigned diagnosis has a female-to-male patient gender ratio is between 0.4 and 0.6. When we restrict the analysis to health conditions that are nearly equally spread across genders, we continue to find an identical pattern of results: male patients consulting female specialists are most likely to obtain a second opinion.

Table 8: Second opinion rate and gender

| | *Dependent variable:* | | | | | |
|---|---|---|---|---|---|---|
| | Probability of second opinion after first consult | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Female specialist | 0.006*** | 0.008*** | 0.006*** | 0.006*** | 0.003*** | 0.008*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.0004) | (0.001) |
| Female patient | 0.003*** | 0.003*** | 0.002*** | 0.002*** | 0.0004 | 0.003*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0002) | (0.0005) |
| Female specialist*Female patient | −0.004*** | −0.005*** | −0.005*** | −0.005*** | −0.003*** | −0.006*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Constant | 0.002 | 0.012* | 0.0002 | 0.005 | 0.001 | 0.019* |
| | (0.004) | (0.005) | (0.004) | (0.005) | (0.003) | (0.010) |
| Specialty FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1,571,075 | 1,585,755 | 1,581,769 | 1,580,248 | 1,559,836 | 766,856 |

Notes: All models are estimated using OLS. Logistic regressions yield similar results.
*p<0.05; **p<0.01; ***p<0.001 (Two-tailed tests).

Table 9: Second opinion rate and gender

| | *Dependent variable:* |
|---|---|
| | Probability of second opinion after first consult |
| Female specialist | 0.007*** |
| | (0.001) |
| Female patient | 0.003*** |
| | (0.0004) |
| Female specialist*Female patient | −0.007*** |
| | (0.001) |
| Constant | −0.014*** |
| | (0.002) |
| Specialty FEs | Yes |
| Year FEs | Yes |
| Observations | 1,093,631 |

Notes: All models are estimated using OLS. Logistic regressions yield similar results.
*p<0.05; **p<0.01; ***p<0.001 (Two-tailed tests).

# Conclusion

In this paper we demonstrate that male and female patients differ substantially when digesting medical advice and determining a course of action to address their health issues. Specifically, while female patients choose to have their medical condition evaluated by a second specialist more frequently than male patients, the decision making process of male patients seem to be driven more by the gender of the first specialist than the decision making process of female patients. Preliminary results highlight the potential consequences of these gendered preferences: if second opinions shift patients from the first specialist to the second, female specialists face relatively high missed earnings compared to male specialists.

# References


Abbott, Andrew. 2014. *The system of professions: An essay on the division of expert labor*. University of Chicago press.

Adams, Tracey L. 2010. "Gender and feminization in health care professions." *Sociology compass* 4:454–465.

Azmat, Ghazala and Rosa Ferrer. 2017. "Gender gaps in performance: Evidence from young lawyers." *Journal of Political Economy* 125:1306–1355.

Azocar, Maria J and Myra Marx Ferree. 2015. "Gendered expertise." *Gender & Society* 29:841–862.

Beckman, Christine M and Damon J Phillips. 2005. "Interorganizational determinants of promotion: Client leadership and the attainment of women attorneys." *American Sociological Review* 70:678–701.

Berger, Joseph, Bernard P Cohen, and Morris Zelditch Jr. 1972. "Status characteristics and social interaction." *American Sociological Review* pp. 241–255.

Blau, Francine D and Lawrence M Kahn. 2013. "Female labor supply: Why is the United States falling behind?" *American Economic Review* 103:251–56.

Blau, Francine D and Lawrence M Kahn. 2017. "The gender wage gap: Extent, trends, and explanations." *Journal of economic literature* 55:789–865.

Bloom, Nicholas, Aprajit Mahajan, David McKenzie, John Roberts, et al. 2020. "Do Management Interventions Last? Evidence from India." *American Economic Journal: Applied Economics* 12:198–219.

Brooks, Alison Wood, Laura Huang, Sarah Wood Kearney, and Fiona E Murray. 2014. "Investors prefer entrepreneurial ventures pitched by attractive men." *Proceedings of the National Academy of Sciences* 111:4427–4431.

Castilla, Emilio J. 2008. "Gender, race, and meritocracy in organizational careers." *American journal of sociology* 113:1479–1526.

Ding, Waverly W, Fiona Murray, and Toby E Stuart. 2013. "From bench to board: Gender differences in university scientists' participation in corporate scientific advisory boards." *Academy of Management Journal* 56:1443–1464.

Eddy, David M. 1984. "Variations in physician practice: the role of uncertainty." *Health affairs* 3:74–89.

Eyal, Gil. 2013. "For a Sociology of Expertise: The Social Origins of the Autism Epidemic." *American Journal of Sociology* 118:863–907. Publisher: The University of Chicago Press.

Fox, Renee C. 1957. "Training for uncertainty." *The student-physician* pp. 207–241.

Freeland, Robert E and Jesse Hoey. 2018. "The structure of deference: Modeling occupational status using affect control theory." *American Sociological Review* 83:243–277.

Freidson, Eliot. 1988. *Profession of medicine: A study of the sociology of applied knowledge*. University of Chicago Press.

Gallotti, Riccardo and Manlio De Domenico. 2019. "Effects of homophily and academic reputation in the nomination and selection of Nobel laureates." *Scientific reports* 9:1–12.

Goldin, Claudia and Cecilia Rouse. 2000. "Orchestrating impartiality: The impact of" blind" auditions on female musicians." *American economic review* 90:715–741.

Greene, Jessica, Judith H Hibbard, and Rebecca M Sacks. 2018. "Does the Race/Ethnicity or Gender of a Physician's Name Impact Patient Selection of the Physician?" *Journal of the National Medical Association* 110:206–211.

Greenfield, Geva, Joseph S Pliskin, Paula Feder-Bubis, Shlomo Wientroub, and Nadav Davidovitch. 2012. "Patient–physician relationships in second opinion encounters–The physicians' perspective." *Social science & medicine* 75:1202–1212.

Gulati, Ranjay and Maxim Sytch. 2008. "Does familiarity breed trust? Revisiting the antecedents of trust." *Managerial and Decision Economics* 29:165–190.

Hall, Judith A, Julie T Irish, Debra L Roter, Carol M Ehrlich, and Lucy H Miller. 1994. "Gender in medical encounters: an analysis of physician and patient communication in a primary care setting." *Health Psychology* 13:384.

Hughes, Everett C. 1963. "Professions." *Daedalus* pp. 655–668.

Ibarra, Herminia. 1992. "Homophily and differential returns: Sex differences in network structure and access in an advertising firm." *Administrative science quarterly* pp. 422–447.

Kanter, Rosabeth Moss. 2008. *Men and women of the corporation: New edition*. Basic books.

Kossinets, Gueorgi and Duncan J Watts. 2009. "Origins of homophily in an evolving social network." *American journal of sociology* 115:405–450.

Mann, Allison and Thomas A DiPrete. 2013. "Trends in gender segregation in the choice of science and engineering majors." *Social science research* 42:1519–1541.

Melamed, David, Christopher W Munn, Leanne Barry, Bradley Montgomery, and Oneya F Okuwobi. 2019. "Status characteristics, implicit bias, and the production of racial inequality." *American Sociological Review* 84:1013–1036.

Moss-Racusin, Corinne A, John F Dovidio, Victoria L Brescoll, Mark J Graham, and Jo Handelsman. 2012. "Science faculty's subtle gender biases favor male students." *Proceedings of the national academy of sciences* 109:16474–16479.

Ody, Christopher, Lucy Msall, Leemore S Dafny, David C Grabowski, and David M Cutler. 2019. "Decreases in readmissions credited to Medicare's program to reduce hospital readmissions have been overstated." *Health Affairs* 38:36–43.

Pedulla, David S and Sarah Thébaud. 2015. "Can we finish the revolution? Gender, work-family ideals, and institutional constraint." *American sociological review* 80:116–139.

Petersen, Trond, Ishak Saporta, and Marc-David L Seidel. 2000. "Offering a job: Meritocracy and social networks." *American Journal of Sociology* 106:763–816.

Podolny, Joel M. 1994. "Market uncertainty and the social character of economic exchange." *Administrative science quarterly* pp. 458–483.

Podolny, Joel M. 2010. *Status signals: A sociological study of market competition*. Princeton University Press.

Prince, Louise A, Lauren Pipas, and Lawrence H Brown. 2006. "Patient perceptions of emergency physicians: the gender gap still exists." *The Journal of emergency medicine* 31:361–364.

Quadlin, Natasha. 2018. "The mark of a woman's record: Gender and academic performance in hiring." *American Sociological Review* 83:331–360.

Ridgeway, Cecilia L and Shelley J Correll. 2004. "Unpacking the gender system: A theoretical perspective on gender beliefs and social relations." *Gender & society* 18:510–531.

Ridgeway, Cecilia L and Kristan Glasgow Erickson. 2000. "Creating and spreading status beliefs." *American Journal of Sociology* 106:579–615.

Ridgeway, Cecilia L and Lynn Smith-Lovin. 1999. "The gender system and interaction." *Annual review of sociology* 25:191–216.

Ruef, Martin, Howard E Aldrich, and Nancy M Carter. 2003. "The structure of founding teams: Homophily, strong ties, and isolation among US entrepreneurs." *American sociological review* pp. 195–222.

Sarsons, Heather. 2017. "Interpreting Signals in the Labor Market: Evidence from Medical Referrals." .

Somaya, Deepak, Ian O Williamson, and Xiaomeng Zhang. 2007. "Combining patent law expertise with R&D for patenting performance." *Organization Science* 18:922–937.

Stuart, Toby E, Ha Hoang, and Ralph C Hybels. 1999. "Interorganizational endorsements and the performance of entrepreneurial ventures." *Administrative science quarterly* 44:315–349.

Zeltzer, Dan. 2020. "Gender homophily in referral networks: Consequences for the medicare physician earnings gap." *American Economic Journal: Applied Economics* 12:169–97.