American Sociological Association

American Sociological Review
2022, Vol. 87(3) 443–477

DOI:10.1177/00031224221087374
journals.sagepub.com/home/asr

SAGE

# Gender in the Markets for Expertise

**Mathijs de Vaan[a] and Toby Stuart[a]**

## Abstract

Stratification in professional careers arises in part from interpersonal dynamics in client-expert dyads. To reduce perceived uncertainty in judgments of the quality of experts, clients may rely on ascriptive characteristics of experts and on pairwise, relational factors to assess the advice they receive. Two such characteristics, expert gender and client-expert gender concordance, may lead to differences in clients' trust in expert advice. To explore these issues, we investigate the incidence of patient-initiated second opinions (SOs) in medicine. In an examination of millions of medical claims in Massachusetts, we find that male patients are much more likely than female patients to obtain an SO if the first specialist they consult is female. Moreover, when the first specialist a patient consults is gender non-concordant and the patient seeks an SO, male patients are substantially more likely to switch to a same-gender specialist in the SO visit. Because patients who lack confidence in the advice of the first-seen specialist infrequently return to this specialist for medical services, female specialists generate lower billings. Analyses of medical spending in follow-up visits suggest that gendered patterns in questioning the advice of medical experts have the potential to contribute substantially to the gender pay gap in medicine.

## Keywords

gender, professions, expertise, trust, wage inequality

There are substantial gender differences in compensation and career advancement rates across the professions (Castilla 2008; Pedulla and Thébaud 2015; Quadlin 2018). These differences persist against a backdrop of shifting occupational choices and structures that have led to a decades-long uptick in the proportion of women in most professions, including law and medicine (Adams 2010; Blau and Kahn 2013, 2017; Mann and DiPrete 2013). Despite near gender equality in rates of entry into the client-based professions—those in which professionals directly interact with clientele—earnings disparities continue to be pervasive (Azmat and Ferrer 2017; Beckman and Phillips 2005; Boulis and Jacobs 2011; Gallotti and De Domenico 2019; Zeltzer 2020).

We argue that the gender gap in professional earnings is partly rooted in client-side, gender-based differences in perceptions of the value of expertise delivered in expert-client dyads. As scholars of the professions observe, expertise is exchanged in relational contexts that are cocooned in core social processes (Azocar and Ferree 2015; Heritage and Maynard 2006; Sandefur 2015). In

[a]University of California-Berkeley

**Corresponding Author:**
Mathijs de Vaan, University of California-Berkeley, 2220 Piedmont Avenue, Room 561, Berkeley, CA 94720, USA
Email: mdevaan@berkeley.edu

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 (1988) 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, the diagnosis-treatment sequence between expert and client is embedded in specific social relationships and cultural understandings (Light 2000; Miller 1994). Given these contextual features, the ritualistic accoutrement orchestrated to create a veil of confidence surrounding these interactions belies frequent ambiguity in how to solve a client's problem, and what constitutes superior guidance (Eddy 1984; Fox 1957; Goodman-Delahunty et al. 2010). This ambiguity paired with asymmetries in knowledge, which occur because clients seek guidance for matters in which they generally are nonexperts, infuses professional practice with client-side uncertainty (Sharma 1997).

If clients must select experts and then act amidst evaluative uncertainty about the quality of advice given, how do they overcome doubts to develop conviction in the guidance they receive? Research shows that in situations marked by uncertainty about quality, evaluators turn to status markers to benchmark their confidence in the quality of a product or service offering (Oldmeadow et al. 2003; Podolny 2010; Ridgeway and Correll 2004; Stuart, Hoang, and Hybels 1999). Moreover, if status and quality are loosely coupled, resources tend to flow to those atop the status hierarchy, sometimes even at the expense of others who provide superior service or advice.

Prior work identifies a host of status markers, such as third-party endorsements (Stuart et al. 1999), racial identity (Melamed et al. 2019), and educational credentials (Rivera 2015). Few status characteristics, however, are as ubiquitous and consequential as gender. Ridgeway and Correll (2004) argue that shared understandings about gender are relatively homogeneous, and expectations about performance differences between men and women in the workplace are deeply ingrained beliefs held by both genders. In this regard, gender is a diffuse status characteristic associated with expectations of competence at almost any task (Webster and Driskell 1983). This makes gender an important factor across many allocative contexts (Abascal and Baldassarri 2015; Brooks et al. 2014). A large body of empirical research supports this argument and demonstrates that both men and women direct fewer resources to women (Brooks et al. 2014; Moss-Racusin et al. 2012). In the uncertainty-infused markets for professional advice, this literature implies that a client's confidence in an expert opinion may depend on the gender of the expert.

Uncertainty in dyadic interactions also may find partial resolution in social similarities between the parties involved. A long research tradition in social psychology demonstrates that individuals' self-concepts consist of unique personal characteristics and their *social identities*: the groups in which they categorize themselves as belonging (Brewer 1979; Tajfel et al. 1971; Tajfel et al. 1979). Scholars have argued that gender is one of the dominant—even ascendant—categories of social identity (Haines, Deaux, and Lofaro 2016; Oakes, Haslam, and Turner 1994), and it is widely understood that common group memberships generate trust (Abrams and Hogg 1990; Cook 2001; Foddy, Platow, and Yamagishi 2009; Hardin 2001; Oldmeadow et al. 2003). This in-group affinity may cushion the possibility that women penalize women in evaluative processes. Therefore, in the professions, the trust a client places in expert advice may not solely depend on the status of the expert but also on the social similarity between members of the client-expert dyad.

We examine these ideas in a study of gender dynamics in physician-patient interactions by studying a proxy for the level of trust in medical advice between gender-concordant versus non-concordant dyads. In our primary analyses, we capture instances in which patients

lack confidence in a physician's advice, as evidenced by their decision to pursue second opinions (SOs). An SO arises when patients seek the council of a second medical specialist concerning a diagnosis, treatment regimen, or prognosis. SOs in expert engagements are more likely to occur when a client lacks confidence in the guidance of the original expert, that is, when the trust the client places in the expert is insufficient. Status-based perceptions of competence make it more likely that both male and female patients will question the guidance of a female expert, but male and female clients may vary in the rate at which they seek SOs following consultations with female experts, because similarity will promote trust in gender-concordant pairs. An important implication of our argument is that if an expert's gender is indeed an antecedent to clients' questioning of advice given, especially if it causes clients to not engage the services of the first opinion (FO) expert, this represents an under-studied mechanism for the gender earnings gap in high-skill professions. Insofar as there is a gendered pattern in the level of confidence in expert advice, this will drive a gender disparity in wage attainment among professionals.

In a dataset comprising all first-time visits to medical specialists for a new-to-the-patient health condition in Massachusetts between 2010 and 2015 (de Vaan and Stuart 2019; Ericson and Starc 2015), we find that patients are more likely to secure an SO after a consultation with a female specialist. In parsing this effect, male patients are much more likely than female patients to obtain an SO if the first specialist they consult is female. Consistent with the theory, the decision of male patients to pursue an SO is more strongly driven by the gender of the FO specialist. These results are robust to many controls, including diagnostic code and primary care physician (PCP) fixed effects.[1] Next, we examine the probability that patients switch to a gender-matched specialist, conditional on pursuing an SO. We find a similar pattern here: when the first specialist a patient sees is the opposite gender, male patients are substantially more likely than female patients to switch to a same-gender specialist for an SO. Finally, we evaluate the implications of these patterns for specialist billings. As expected, we find that female specialists generate 10.7 percent lower billings than do their male counterparts in the year following a first visit for observably equivalent patients. In line with our findings on the incidence of SOs, this gender gap is driven much more by the healthcare consumption choices of male patients than of female patients.

## THEORY

When encountering problems or decisions that require specialized knowledge, clients consult experts to select a course of action. 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 nonexpert 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.

Because the expert generally has far deeper knowledge than the client, the professions embody decision-making under asymmetric information (Bloom, Standing, and Lloyd 2008; Sharma 1997). In fact, many scholars regard restricted access to esoteric knowledge as the defining 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 1988). 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 learning (Abbott 1988; Freidson 1988).

The asymmetry in knowledge between clients and experts is the contextual feature that elevates the need for trust in client-expert dyads. As Hughes (1963:656) 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." Restating Hughes: a client generally can neither ex ante evaluate an expert's skill nor infer the quality of advice from ex post outcomes. Extending this uncertainty even further, scholars in symbolic interactionism question whether expertise itself can ever be truly objective or fully codified. Rather, because expertise is situated and shared in embedded contexts and is characterized by myriad evaluative uncertainties, the nuances in social interaction are part and parcel of it. This contrasts with a view of expertise as an objective resource that is created and controlled by the professions and dispensed in a more regularized, universalistic manner (e.g., Barley 1989; Becker 1970; Hughes 1963; Parsons 1951). 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.

Given the rampant uncertainty in client-expert dyads, a pertinent literature documents how individuals respond to high levels of uncertainty in decision settings. In product markets and cultural arenas, actors' evaluations of goods of uncertain quality often default to assessments of the identity and characteristics of producers, because the latter is easier to size up than the former (Azoulay, Stuart, and Wang 2014; Correll et al. 2017). We anticipate this also to be true in evaluations of experts. How do we assess the quality of a thoracic surgeon when we, the patient, know nothing of the matter? In such situations, we evaluate more concrete, easier-to-observe signals of a provider's quality, such as the names on the framed degrees adorning the office walls during consultations, the stature of the organization that employs the expert, or the referrals we receive from others who we assume to be better informed than we are. And often, the literature shows, we make assessments based on the gender of the expert (Ridgeway and Smith-Lovin 1999).

Ascriptive characteristics—including gender—frequently influence how people judge ability. This is true even in contexts without an actual relationship between these attributes and merit (Berger, Cohen, and Zelditch 1972; Ridgeway and Erickson 2000). For instance, it occurs 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 (Castilla 2008; Goldin and Rouse 2000; Petersen, Saporta, and Seidel 2000). Using gender to infer competence is not the exclusive province of male evaluators but is prevalent among female assessors as well. As Ridgeway and Correll (2004:513) explain, "these abstracted, hegemonic understandings of men and women are roughly consensual in that virtually everyone in the society knows what they are and likely expects that most others hold these beliefs." As one example, in academe, a recent study of applicants to positions for associate and full professorships found that male academics were more likely to be promoted and that female evaluators were no more favorable toward female candidates than were males (Bagues, Sylos-Labini, and Zinovyeva 2017).

Notably, gender as a status marker may be reinforced by behaviors that are consistent with the perceived status differentials between men and women. For example, recent literature highlights the male-female heterogeneity in the portrayal of confidence when individuals convey their opinions. This work suggests that assertive and self-promotional communication styles are perceived as masculine traits, and that it is counter-normative for women to exhibit these characteristics (Rudman and Phelan 2008; Williams and Tiedens 2016). This work follows research in social psychology on advice-taking, which consistently finds that experts' portrayal of

confidence increases the trust others place in them and the likelihood that individuals will follow the advice given (Abascal and Baldassarri 2015; Sniezek and Van Swol 2001). If male experts express more confidence in the advice they provide to their clients, and if this self-assurance in communication affects clients' decision-making, the resulting behavioral patterns likely reinforce those that occur due to the allocative implications of a gender-based status hierarchy.

The argument to this point is that an expert's gender will serve as a status marker that signals quality in the uncertainty-laden markets for expertise and that, on the client-side, this may occur regardless of client gender. Alongside general perceptions of expertise, however, interpersonal interactions between clients and experts are routine in the provision of professional services. Moreover, the nature of the client-expert interaction can be profoundly personal. This shifts interaction from arm's length to engaged. It is therefore important to consider how, in the absence of a history of working together, dyad-level characteristics may facilitate trust, and by extension, the resolution of doubt about expert quality.

A rich line of work in social psychology demonstrates that salient sociodemographic characteristics, among which gender is paramount, shape social identities and promote feelings of belonging when interacting with other identity-group members (Abrams and Hogg 1990; Tajfel et al. 1971; Tajfel et al. 1979). This in-group orientation promotes a variety of behaviors, such as abiding by common norms (Terry and Hogg 1996), holding similar cultural understandings (Chiu et al. 2000), and communicating with greater candor (Prisbell and Andersen 1980). Sharing a prominent sociodemographic characteristic with an alter may also generate higher levels of trust compared to interactions with out-group members (Foddy et al. 2009; Oldmeadow et al. 2003).

Brewer (2008) presents several mechanisms that may explain why shared social identities promote trust. These include heuristic thinking, which is a mental shortcut that entails placing trust in in-group members; social projection, which reflects one's own sense of trustworthiness onto in-group peers who resemble oneself; and social identification, which is placing trust in in-group peers as a means to benefit the group. These mechanisms are not mutually exclusive, and the strength of each is likely to be context-specific (Evans and Krueger 2009). However, regardless of the mechanism responsible for linking social identity and trust, gender concordance in expert-client dyads is expected to increase trust and thereby reduce a client's tendency to question a provider's expertise.

These findings on in-group membership and trust are echoed in sociological research on homophily (Cook 2001; Melamed et al. 2020). For example, Rivera (2012) finds that in a hiring context, cultural matching between candidates and evaluators facilitates trust. Likewise, in interpreting their finding that law firms are more likely to promote female partners when their corporate clients are women-led, Beckman and Phillips (2005:682) speculate that "women-led corporate clients may have greater trust for and be comfortable communicating with women attorneys."

Status-induced trust and similarity-induced trust likely coalesce to influence perceptions of male and female experts. The status-based argument leads to the prediction that clients, regardless of gender, are more prone to question the expertise of female experts. Similarity-based accounts suggest that confidence in expert advice may be greater in gender-concordant pairings. The net effect of these two social dynamics depends on their strength. For example, if the status-based effect and the similarity-based effect for female clients are equally strong, female clients should exhibit little or no differences in behavior after consulting a female versus male expert. Male clients, on the other hand, should display more divergent behavior depending on the gender of the expert they consult. Specifically, male clients should be substantially more likely to doubt the guidance of female experts

compared to the advice given by an otherwise comparable male. Therefore, in aggregate, *there will be less trust in the recommendations of female experts, and male clients will evince a stronger inclination to question the recommendations of female experts*. These arguments lead us to two propositions about gender dynamics in markets for expert opinions:

> *Proposition 1:* Clients will be more likely to question the recommendations of a female expert than the recommendations of a comparable male expert.
>
> *Proposition 2:* Compared to female clients, male clients will be more likely to question the recommendations of a female expert relative to a male expert.

Before introducing our empirical context, it is important to highlight four scope conditions that bound these propositions. These factors are nonbinary; the more each is present in a professional setting, the more likely it is that the behavior described in our propositions will unfold. The first scope condition is that there is significant uncertainty in assessing the quality of experts and the advice they give. Typically, this occurs in client-expert dyads because the patron is nonexpert and therefore lacks the knowledge to fully understand situational diagnoses and remedial options. The second scope condition is the presence of generally held, gendered stereotypes for the competence of experts in the given professional field. The greater the extent of occupational sex-typing, the more salient gender will be as a status marker that can resolve the aforementioned uncertainty. Third, the presence of an involved or engaged interaction is a scope condition because it amplifies the salience of social identities in the formation of interpersonal trust. In- and out-group distinctions loom largest in contexts in which interactions are personal and are socially, economically, or professionally consequential. Finally, status markers and social similarity are especially likely to influence trust in *first-time* interactions (Dahlander and McFarland 2013). Prior research suggests that repeated exchange between actors promotes trust (Bian 1997; Cook and Emerson 1978; Kollock 1994) and therefore attenuates the reliance on identity signals, such as status and social similarity.

We believe these are the main scope conditions that contain the purchase of the two propositions. In considering professional settings in which these boundary conditions are met, many stand out. Examples include attorney-client interactions (Howieson and Rogers 2019), economists providing legal consulting (Del Rossi and Hersch 2020), and financial advisors providing investment recommendations (Chiou and Droge 2006). As we describe in detail in the following section, they also apply to physicians providing medical care.

## EMPIRICAL CONTEXT

The profession we study is medicine, and we focus on interactions between medical specialists (experts) and patients (clients). This is a compelling context to explore the two propositions because, as we will describe, it is characterized by high uncertainty regarding the quality of service provision, there is a large literature on gender stereotyping and sex-typed roles and behaviors in medicine, interactions in patient-physician dyads are typically engaged and consequential, and in contrast to primary care physician (PCP) visits, for new-to-the-patient health conditions, patients and medical specialists almost never have a prior working relationship. Thus, the context satisfies the boundary conditions of the theory.

When patients develop a non-emergency health condition, they often begin with a visit to a PCP. If the PCP determines that the condition requires a specialist, the patient is provided with a referral. The choice of a specific specialist often is a joint function of the PCP's recommendation and the patient's preference (Song, Sequist, and Barnett 2014). Following a referral, the patient typically schedules an office visit in which the specialist performs an evaluation, communicates a diagnosis or a

need for additional testing, and suggests treatment options. The patient must then decide whether to adopt the advice given. If patients lack conviction in the advice, they may either seek an SO from another qualified specialist or simply decide not to return for a follow-up visit with the first specialist.[2] As we will describe in much greater detail, we are able to develop a proxy for a patient's level of confidence in a physician's opinion by investigating the incidence of *patient-initiated* second opinions (SOs).[3] Building on the idea that SOs often manifest patients' lack of confidence in the medical advice received from a specialist provider, in this section we further develop and refine the two general propositions to the specific case of the incidence of SOs in medicine.

The subject of trust in the patient-physician interaction is a salient feature of much of the sociological research on this dyad. For example, some work proposes that "patient-centered" medicine, which promotes more collaborative patient-physician interactions, may help reduce the knowledge asymmetry between patients and physicians. In their review of this literature, however, Pilnick and Dingwall (2011) find that the evidence is mixed at best, suggesting that knowledge asymmetries remain and that the need for trust is a persistent feature of patient-physician interactions.[4] Consistent with this view, Greenfield and colleagues (2012:1203) offer a textured description of the patient-physician interaction: "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" (see also Heritage and Maynard 2006).

Alongside the literature on trust, extensive work examines gender dynamics in medicine in general and in physician-patient interactions in particular (Boulis and Jacobs 2011; Pilnick, Hindmarsh, and Gill 2009; Timmermans 2020). Some of this work highlights the role of physician gender in making inferences about ability or quality. In a telling study, Prince, Pipas, and Brown (2006) surveyed patients in a hospital emergency department (ED) to ask whether they had been visited by a physician, when *all* patients had received a consultation. In 93.3 percent of consultations with male doctors, patients recognized the interaction as a physician visit, compared to just 78.5 percent of consultations if the physician was female. In short, female physicians were less likely than male physicians to be recognized as doctors. This pattern of missed recognition was similar for male and female patients, mimicking the public's gendered perception of an ED physician.

Likewise, Sarsons (2017) finds that referring physicians interpret patient outcomes differently depending on the gender of the specialist provider. She shows that 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. This pattern was similar among male and female referrers. The upshot of this body of work is that gender stereotypes are rampant in medicine, and perceptions of gender-based status and quality differences are widely held. These factors pose routine challenges for female physicians (Koven 2017). They also imply that patients are likely to rely on a specialist's gender to draw inferences about the quality of the medical advice they receive. We therefore hypothesize the following:

> *Hypothesis 1*: Patients will be more likely to seek a second opinion after consulting a female specialist than after consulting an otherwise comparable male specialist.

Shifting our focus to patient-physician pairs as a specific instance of client-expert dyadic exchange, does gender concordance in these dyads contribute to the emergence of trust? To begin, note that patients frequently select same-gender primary care physicians (Fang, McCarthy, and Singer 2004), with as many as 78 percent of female patients choosing female PCPs. (As we will demonstrate,

our data also show strong evidence of gender homophily in specialist selection.) Presumably, part of the explanation for patients' preference for a same-gender physician is the anticipation of greater trust in the relationship. In a study of audiotaped patient-clinician visits, Meeuwesen, Schaap, and Van Der Staak (1991) find that male general practitioners spend more time interviewing female than male patients, perhaps because of an absence of trust needed to converse candidly. Consistent with these results, Cook and colleagues (2004) find that patient-physician gender concordance facilitates trust by creating more transparent relationships.

A large literature explores gender concordance and communication patterns in patient-physician interactions. Sandhu and colleagues (2009) observe higher levels of communication and greater focus on the patient in same-gender patient-physician dyads, particularly female-female pairs. Moreover, the gender composition of the dyad influences communication styles, nonverbal communication, demonstrations of power, and consultation length. In an early study of gendered communication patterns in physician-patient interactions, West (1984) noted that interruption of speech is a conversational manifestation of dominance in doctor-patient dyads. Although the number of observations in the study is small, the results suggest that male physicians interrupt patients in assertions of authority, whereas female physicians often are interrupted, particularly by male patients. Perhaps because of the enhanced degrees of communication and trust, physicians report greater comfort providing preventative care for same-gender patients (Lurie et al. 1998).

If gender concordance facilitates interpersonal trust in patient-physician dyads, it also may moderate the effect of gender-based status distinctions among physicians. Specifically, for male patients, the absence of confidence in female specialists will likely widen the gap in the incidence of SOs between female and male specialists. For female patients, the higher anticipated trust in female specialists will attenuate the gap in SO-seeking between male and female specialists. Phrasing this formally, we hypothesize the following:

> *Hypothesis 2:* Compared to female patients, male patients will be more likely to seek a second opinion following a consultation with a female specialist relative to a male specialist.

## DATA AND METHODS

For this study, we examined the Massachusetts All Payers Claims Database (MA APCD). The MA APCD is collected and curated by the Center for Health Information and Analysis (CHIA) and contains remarkably comprehensive information derived from the medical claims of virtually every resident in Massachusetts. Our analyses rely on version 4 of these data, which contains observations between January 1, 2010, and December 31, 2014.

Massachusetts requires all health insurers in the state to report to CHIA the details of every medical claim they process. CHIA collects these data and prepares them for use in research. For instance, CHIA created a hashed identifier to link the medical records of individuals who change insurance plans over time. For this study, we principally draw from the Medical Claims data file, which contains more than 650 million distinct medical claims. These data include patient and physician identifiers, diagnostic codes, dates and locations of consultations, medical procedures performed, dollar amounts charged, and, importantly, in the case of specialist visits, 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 Strategy and Identifying Second Opinions*

Data as comprehensive as the MA All Payers Claims Database have only become available recently, and there is no widely agreed-upon method for identifying patient-initiated SOs

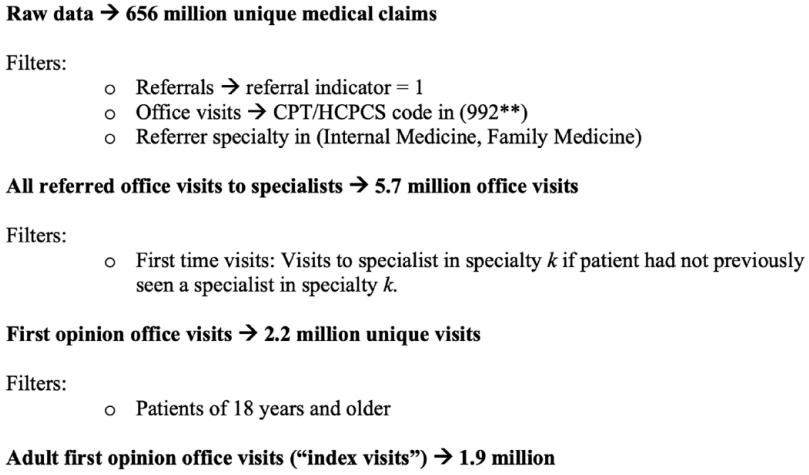


**Figure 1.** Sampling Steps
*Note:* The referral indicator in the MA APCD indicates whether a medical visit to a physician was preceded by a referral from another physician. Because specialists may also refer to one another, we limit referrers to PCPs (internal medicine and family medicine) to maintain our focus on *patient-initiated* second opinions.

in these data. The small body of research on SOs is mostly survey-based, and a comprehensive literature review by Payne and colleagues (2014) found only 13 articles that met their inclusion criteria. In a more recent study, Shmueli, Shmueli, and colleagues (2019:3) note, "To the best of our knowledge, all studies that evaluated SO utilization so far were based on patient self-reported surveys and not on objective data." In response to this observation, Shmueli, Shmueli, and colleagues (2019) introduce a strategy to identify SO visits to medical specialists in a sample of first opinion (FO) visits. Specifically, they define an SO to occur when the patient visits a second specialist in the same specialty as an FO specialist within a three-month period following the FO visit. We broadly follow this strategy, but we take several additional steps to construct a sample of FO visits in which patients have discretion to seek an SO. We use this sample as the census of medical consultations that may be followed by an SO.

We begin by extracting all 4.6 million office visits to specialists that resulted from a referral by a PCP.[5] We next limit this sample to only first-time office visits to a physician in specialty $k$ conditional on the patient having not previously seen a provider in specialty $k$, $t < T$. Doing so reduces the sample to about 2.2 million observations. Finally, we limit the sample to patients 18 years and older, which yields an analysis sample of 1,887,253 "index" visits, following which a patient might initiate an SO. For ease of replication, Figure 1 describes the steps to create the final sample.

Implementing these sampling steps achieves several goals. First, by focusing only on in-office consultations, we eliminate specialties such as radiology and pathology in which patients and experts rarely meet in person, and therefore gender dynamics are unlikely to come into play. Second, by including only office visits resulting from a referral by a PCP, we exclude cases that begin with an emergency department visit where, due to the time sensitivity of the health condition, patients may have limited scope to seek an SO. Third, by removing patients younger than 18, we exclude cases where the patient may not be the primary decision-maker.[6] Finally, by focusing on new-to-the-patient health conditions for which the patient has no observed medical history in specialty $k$, we achieve two goals: (1) we reduce the likelihood patients

and specialists have preexisting relationships, and (2) we sample FO consultations in which patients are unlikely to have developed deep expertise in their condition. These screens help align the analysis sample to the boundary conditions of the theory: gender effects will be most prevalent in de novo relationships and in situations of higher knowledge asymmetry.

In the sample of index visits, the next step is to identify visits in which the patient sought an SO. Here, we follow Shmueli, Davidovitch, and colleagues (2019) and identify all FOs in which a patient sees another specialist in the same specialty as the FO specialist within a limited time frame. In contrast to Shmueli, Davidovitch, and colleagues (2019), we extend the window of the second consultation to six months, because appointment wait times often extend beyond three months for non-emergency health conditions (Flemons et al. 2004; Neimanis et al. 2017; Olayiwola et al. 2016). However, sensitivity analyses show that our findings are robust to using a three- and a one-month window.[7]

In summary, we define an SO to 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 PCP. About 4 percent of all FO visits are followed by a so-defined SO.[8] This primary outcome variable is binary and equals one if a patient obtains an SO.

Because the number of studies of SOs in medical claims is limited, we conduct several analyses to validate the outcome measure. The detailed results of these analyses are described in Appendix A. First, the existing, survey-based literature suggests that SOs should be more prevalent for serious health conditions. To evaluate whether this relationship describes our measure of SOs, we compute three measures of severity of the diagnoses made in each of the 1.9 million index visits. We do so by calculating the average one-year medical spending, the average one-year hospitalization probability, and the average one-year surgery probability *associated with every diagnosis code in the claims data*.[9] We then link these averages to the diagnoses in the index visits and test whether index visits followed by an SO have a higher mean for each of these severity measures. As shown in Appendix Figure A1, we find that they do, and that the difference is meaningful and statistically significant.

In a second test, we address the concern that SOs may occur because of a mismatch between the expertise of the FO specialist and the expertise needed to successfully treat a patient's health condition. Medical fields can have elaborate sub-specializations, and for this reason or because a PCP's initial diagnosis turns out to be incorrect, patients occasionally may be mis-assigned to specialists in the referral process. To assess this concern, we extract treatment histories of specialists consulted in the FO visit to compare them to those of the specialist in the SO visit. To create a benchmark that can be used to evaluate the similarity in treatment histories of the FO and SO specialists, we randomly sample counterfactual SO specialists who are in the same specialty as the FO specialist. The intuition for this analysis is that, if patients are frequently referred to specialists who do not have the expertise to treat their conditions, the treatment experience of the FO and SO specialists should be more dissimilar to one another than would be true for the randomly matched, benchmark pair. Results of this analysis (see Appendix Figure A2) show that, if anything, the expertise of the FO and SO specialists is considerably *more similar* than the expertise of the FO specialist and a randomly selected, same-specialty clinician.

Finally, patients may develop a new health condition between the first and second specialist visits. If this health condition also requires medical care from a specialist in the same specialty as the FO specialist, we would mistakenly label the second visit as an SO. We examine the frequency of such a scenario and find it rarely occurs. In summary, these validation exercises bolster our confidence in the procedure we use to identify SOs, but to err on the side of caution, we conduct

several robustness checks in which we drop cases that may be considered ambiguous; the results are robust to these deletions.

### *Research Design and Analytic Roadmap*

In our main analyses, we estimate the probability that a patient will seek an SO following an index visit as a function of the patient's and specialist's gender. Specifically, we estimate the coefficients of a regression model that includes patient gender, specialist gender, an interaction of the two variables, and an extensive vector of control variables. The fully saturated regression equation is shown in Equation 1:

$$\begin{aligned} \mathrm{P}(\mathrm{SO}=1) = {} & \beta 1 * \textit{Female Specialist}_i \\ & + \beta 2 * \textit{Female Patient}_k \\ & + \beta 3 * \textit{Female Specialist} \\ & * \textit{Female Patient}_{ik} \\ & + \beta 4 * X_{ik} + \epsilon \end{aligned} \tag{1}$$

where $i$ refers to the patient, $k$ refers to the specialist, and $X_{ik}$ is a vector of control variables.

We are interested in two empirical estimands (Lundberg, Johnson, and Stewart 2021). First, in a regression model that does not include an interaction between *female patient* and *female specialist*, we are interested in the coefficient for *female specialist*. This coefficient will capture the conditional mean of the difference in which patients of female specialists and patients of male specialists seek SOs. This quantity directly maps onto our first hypothesis. The second empirical estimand we are interested in is given by the difference in $\beta_1$ (i.e., the rate difference at which male patients seek SOs after seeing a female versus male specialist) and the sum of $\beta_1$ and $\beta_3$ (i.e., the rate difference at which female patients seek SOs after seeing a female versus male specialist). This quantity directly maps onto our second hypothesis.

A primary threat to identifying whether patients' decisions to seek an SO are driven by the gender of the specialist is the possibility of other, systematic differences between male and female specialists. In other words, there may be one or more confounders, which are unobserved differences that are correlated with specialist gender and the probability of obtaining an SO. Perhaps the most likely confounders include unobserved specialist quality, risk perceptions of male versus female specialists, and the severity of patients' health statuses.[10] Consider a scenario in which female specialists are of lower quality than male specialists: they provide less accurate diagnoses. Such a difference should lead patients to have less trust in the recommendations of female specialists and higher rates of SOs, but our goal is to identify trust differences that are identity based, not merit based. This highlights the salience of including control variables that capture quality differences between specialists. A similar logic applies to differences in risk perceptions between male and female specialists. If female specialists consistently diagnose medical conditions as carrying more risk, patients will be more likely to seek SOs. And finally, if patients with more severe medical conditions seek out female specialists, the SO rates will be higher for female specialists even if the trust patients have in the advice does not vary with specialist gender. (However, based on research on gender as a status marker, one would likely predict the opposite: less healthy patients will be more likely to select male specialists because of a perception of greater competence.)

Our first strategy to address these issues is to include a set of direct controls, which we describe in detail below. Although including these controls may not meet the assumption that, conditional on the covariate vector, $corr(\text{female specialist}_i, u_i) = 0$, their inclusion allows us to implement a second approach. Because we observe how much the main estimate attenuates with the inclusion of controls, we can build on Altonji, Elder, and Taber's (2005) method for using information about coefficient stability to quantify how large of an impact omitted variables would need

to have for the estimand of interest to be reduced to zero. We use a recent enhancement to this method to assess the likelihood that unobserved confounders might invalidate our findings (Oster 2019). Our third strategy is to estimate models with specialist fixed effects (FEs). To implement the FE estimator, we run two separate regressions, one for female and one for male specialists.[11] These specifications allow us to ascertain whether male and female patients request SOs from the *same specialist* at different rates. In other words, all unobservable time-stationary characteristics of the physician, including quality, are held constant in these regressions.[12]

Finally, in testing the second hypothesis, unobserved quality differences between male and female specialists should be less of a concern. Note that *if* the gap between the SO rate for the patients of male and female specialists is driven by unobserved specialist differences, both male and female patients are likely to respond to these differences; there should not be a heterogeneous response by patient gender. In other words, our strategy to identify the effect described in Hypothesis 2 relies on the relative difference between male and female specialists *by patient gender*, which effectively accounts for unobserved differences between specialists (Zeltzer 2020).

In addition to our main analysis, we conduct three supplemental analyses to aid in the interpretation of the main findings. First, in the development of our argument we have been silent about how patients select the gender of FO specialists. However, understanding the pattern of gender sorting into FO consultations is important for interpreting the results. If patients tend to choose gender-matched specialists, we can assume the average patient that matches to an opposite-gender specialist for an FO consultation to be likely to hold at most a weak preference for a same-gender clinician. If a gender-concordant sorting process drives the sample of FO visits, which then becomes the set of index consultations for SOs, estimates from our main models will likely be driven toward zero compared to a setting in which patients (and clients, more generally) have less control over the selection of the specialist. To provide insight into this first-stage sorting process, we estimate the incidence of gender homophily in the matching process for FO specialists. We describe this analysis in more detail later.

Second, we examine the role of gender concordance in switching patterns observed for patients deciding to seek an SO. We refer to these choices as intensive margin decisions, because we only observe these decisions for patients who seek an SO. If gender (and the associated variation in confidence in a recommendation) is an important factor for patients to seek an SO, one would expect gendered switching patterns to align with those observed in our main analysis. In other words, male patients seeing female physicians for an FO consultation should be particularly likely to switch to a male doctor for an SO. We therefore see this analysis as a validation check for our main analysis. The regression equation for these estimates can be written as follows:

$$\begin{aligned} &\mathrm{P}(\text{Gender switch} = 1) \\ &\quad = \beta 1 * \textit{Female Patient}_i \\ &\quad + \beta 2 * \textit{Female Specialist}_k \\ &\quad + \beta 3 * \textit{Female Patient} \\ &\quad * \textit{Female Specialist}_{ik} \\ &\quad + \beta 4 * X_{ik} + \varepsilon \end{aligned} \tag{2}$$

Third, we evaluate the financial impact of the theory. As we will demonstrate in the next section, patients who lack confidence in a specialist are not only more likely to pursue an SO, but they are also less likely to schedule a follow-up consultation with that physician. These behaviors reduce the patient's consumption of medical services from the FO specialist. If the hypotheses are supported, the implication is that female FO specialists will have fewer opportunities to generate income, particularly from male patients. To examine this, using the sample of FO visits, we estimate regressions of the medical spending *of each patient in our sample* on services provided by the FO specialist in the

year following the index visit. The regression equation is as follows:

$$\begin{aligned}\text{Log}\left(1\text{ year spending}\right)_{ik} &= \beta 1 * \textit{Female Patient}_i \\ &+ \beta 2 * \textit{Female Specialist}_k \\ &+ \beta 3 * \textit{Female Patient} \\ &* \textit{Female Specialist}_{ik} \\ &+ \beta 4 * X_{ik} + \varepsilon\end{aligned} \tag{3}$$

## *Explanatory Variables and Descriptive Statistics*

For each index visit in our sample, we extract and construct a set of variables, including the outcome variable, main explanatory variables (*patient gender* and *specialist gender*), and variables that may confound the relationship between specialist gender and the probability of seeking an SO. We first introduce our main variables followed by a rationale for inclusion in our regression models.

Patient and specialist gender are two binary variables that equal one for females (i.e., male is the reference category [Johfre and Freese 2021]). The first control variable, *patient age*, is measured at the time of the visit. Because the association between patient age and the probability of an SO is unlikely to be linear, we include patient age as a spline (i.e., 18 to 44, 45 to 54, 55 to 64, and 65+). We also construct the *Charlson comorbidity score* for each patient in the sample based on their medical history in the year prior to the index visit. This is a standard co-morbidity index that is predictive of a patient's one-year mortality risk. The score is based on the presence of 22 serious health conditions, including heart disease, AIDS, and cancer. Each health condition is assigned a score of 1, 2, 3, or 6, depending on the associated mortality risk. The overall score is a sum across the 22 conditions. Another patient-level control variable is *insurance type*. We include this variable as a set of dummies that includes health maintenance organization (HMO), point of service (POS), preferred provider organization (PPO), Medicaid, exclusive provider organization (EPO), and other insurance type.

Next, we include a set of controls for each FO specialist. We include specialists' medical school *graduation year* as a three-piece spline: pre-1980, 1980 to 2000, and 2000 and later. *Specialist experience with condition* measures the relative focus of the specialist on the patient's specific medical condition. For the diagnosis code assigned to the patient in the FO visit, we assess specialists' relative experience with that diagnostic code in their full practice history in the year prior to the index visit.[13] *Specialist top-15 medical school* is a dummy variable equal to one if the specialist graduated from a top-15 medical school, based on *U.S. News* rankings.[14] We also include *fixed effects for the specialties* of the specialists seen in the first visit.

Finally, we include *proportion female specialists in HRR*, *female PCP*, and a full suite of *diagnosis fixed effects*. Proportion female specialists in HRR is the proportion of women in the FO specialty in the month prior to the FO visit in the specialist's "hospital referral region," which is a geographically delineated area based on referral patterns for tertiary medical care. HRRs are created and published by the Dartmouth Atlas of Health Care. Female PCP is a dummy that equals one if the referring PCP is female. Finally, the diagnosis code fixed effects are based on the 925 distinct three-digit ICD-9 diagnoses in the sample. ICD-9 is a hierarchical classification system with a maximum of five digits, but diagnoses are often coded with three, four, or five digits. For example, physicians may diagnose a patient with migraine (346), migraine with aura (346.0), or migraine with aura, with intractable migraine, so stated, without mention of status migrainosus (346.01). We include the fixed effects at the three-digit level, which prevents nested diagnoses from being represented as completely independent of one another. In addition, using five-digit codes results in many cells with a single observation.

The purpose of including these covariates is to adjust for potential confounding factors. For example, we include patient controls such as the *Charlson comorbidity score* to account for matching of patients to specialists

of a specific gender based on health status. As mentioned before, if patients with more serious illnesses are more likely to select a female FO specialist, it would be unsurprising if the patients of female specialists seek more SOs. Including a rich set of controls to adjust for the severity of health conditions reduces the likelihood that we spuriously interpret the link between specialist gender and seeking an SO as one driven by a patient's lack of confidence in the expert's opinion.

The physician controls we include adjust for observable quality or risk-perception differences between male and female specialists. For example, *specialist graduation year*, *specialist experience with condition*, and *specialist top-15 medical school* may capture different dimensions of specialist quality, and may also be correlated with specialist gender and with the SO rate. Likewise, if the risk perception of conditions varies between male and female specialists, they are likely to express this through the diagnosis code used to describe the patient's condition (Landon and Mechanic 2017). The goal of including diagnosis fixed effects is to adjust our estimates for such differences. Finally, the inclusion of medical specialty fixed effects accounts for the sorting of male and female physicians into specialties with higher or lower SO base rates. These condition out all differences that are attributable to gender selection into specialties.

Table 1 shows descriptive statistics for the sample, broken out by the gender of the specialist in the index visits. Several differences between male and female specialists are apparent. The first row of the table shows that patients of female specialists obtain SOs more often than do patients of male specialists. Suggestive of a financial penalty for female specialists, the second row of the table shows that subsequent-year per-patient billings for male specialists are approximately 25 percent higher than for female specialists. Next, the female patient and female PCP variables suggest that gender homophily drives patients' choice of specialist providers. Female specialists see slightly younger patients; this occurs because female specialists have, on average, entered the workforce more recently than male specialists (see specialist graduation year), and there is age homophily between patients and providers. The Charlson score is slightly higher for the patients of male specialists, a difference that is partly due to gender sorting of physicians into specialties and the slightly older age of patients seen by male specialists. Finally, patients seen by male and female specialists have largely similar healthcare insurance coverage.

Turning to characteristics of the specialist, male specialists graduated earlier than female specialists, they have slightly less experience with the medical conditions they diagnose, and they are less likely to have graduated from a top-15 medical school. In the detailed breakdown of specialty by gender, male and female clinicians specialize in different fields. For example, women are over-represented in dermatology, and men are over-represented in orthopedic surgery. Finally, the proportion female specialists in HRR is higher for women than for men, suggesting some spatial clustering of male and female specialists.

Table 2 presents descriptive statistics based on whether an index visit resulted in an SO. The first row confirms our previous speculation: patients who seek SOs are much less likely to return to the FO specialist for subsequent care, which greatly reduces the average one-year patient spending on services from that physician. The table also shows that patients who seek SOs are less healthy, as indicated by a higher Charlson score, but they are otherwise relatively similar. In line with Table 1, Table 2 shows the patients of female specialists comprise a higher fraction of the SO sample cases than the non-SO cases. Also, in the subset of cases that lead to an SO, the FO specialist is more likely to have graduated from a top-15 medical school. This may seem counterintuitive, but it occurs because specialists from elite medical schools are more likely to practice at academic medical centers, where specialists see patients with more complex health conditions. Finally, some medical specialties, such as orthopedic surgery, are over-represented in SO cases and others are under-represented. These differences are

**Table 1**. Descriptive Statistics by Gender of the Specialist in the Index Visit

| | Male Specialist | Female Specialist |
|---|---|---|
| *Outcome Variables* | | |
| Second opinion | .03 (.18) | .04 (.19) |
| One-year spending (in dollars) | 565 (2559) | 456 (2520) |
| *Patient Characteristics* | | |
| Female patient | .55 (.50) | .72 (.45) |
| Patient age (Mean) | 48.61 (13.90) | 46.42 (13.92) |
| Charlson index score (Mean) | .61 (1.08) | .54 (1.00) |
| Insurance type | | |
| Health maintenance organization (HMO) | .73 (.45) | .73 (.44) |
| Point of service (POS) | .08 (.27) | .09 (.28) |
| Preferred provider organization (PPO) | .07 (.25) | .08 (.27) |
| Medicaid | .03 (.17) | .03 (.16) |
| Exclusive provider organization (EPO) | .02 (.13) | .02 (.13) |
| Other insurance type | .08 (.27) | .06 (.24) |
| *Specialist Characteristics* | | |
| Specialist graduation year (mean) | 1986 (11) | 1993 (9) |
| Specialist experience with condition (mean) | .03 (.08) | .04 (.08) |
| Specialist top-15 medical school | .18 (.39) | .20 (.40) |
| Provider specialty | | |
| Dermatology | .14 (.34) | .29 (.45) |
| Orthopedic surgery | .14 (.34) | .03 (.16) |
| Otolaryngology | .06 (.24) | .04 (.21) |
| Urology | .07 (.25) | .01 (.12) |
| Surgery | .06 (.23) | .06 (.23) |
| Gastroenterology | .06 (.24) | .04 (.20) |
| Ophthalmology | .05 (.22) | .04 (.20) |
| Neurology | .04 (.20) | .05 (.22) |
| Cardiovascular disease | .05 (.22) | .02 (.15) |
| Obstetrics and gynecology | .01 (.12) | .08 (.27) |
| Other specialty | .31 (.46) | .33 (.47) |
| *Other Variables* | | |
| Female PCP | .40 (.49) | .54 (.50) |
| Proportion female specialists in HRR (mean) | .24 (.15) | .37 (.16) |
| Observations | 1,427,106 | 460,147 |

*Note:* This sample includes index visits for adult patients who were referred to a specialist by a PCP. Calendar year and diagnosis codes are omitted for brevity. The values in parentheses are standard deviations from the mean. The male-female differences in all variables in this table are statistically significant based on *t*-tests and chi-square tests.

at least partly driven by variation in the complexity and severity of health conditions treated by different specialties.

## *Evaluating Baseline Homophily*

To determine whether specialist gender is a factor in selecting a specialist, we estimate dyadic choice models of the match probabilities between a patient and all possible FO specialists. Both the patient and the PCP are likely to influence specialist selection, so we account for both patient-specialist and PCP-specialist gender concordances. To model the probability that patient $i$ selects focal specialist $k$, we estimate conditional logistic regressions, which is the standard approach to estimate discrete choice models

**Table 2**. Descriptive Statistics by Whether the Index Visit Results in a Second Opinion

| | Second Opinion | No Second Opinion |
|---|---|---|
| *Outcome Variable* | | |
| One-year spending (in dollars) | 369 (2,286) | 545 (2,559) |
| *Patient Characteristics* | | |
| Female patient | .60 (.49) | .59 (.49) |
| Patient age (Mean) | 47.58 (14.26) | 48.10 (13.92) |
| Charlson index score (Mean) | .67 (1.16) | .59 (1.06) |
| Insurance type | | |
| Health maintenance organization (HMO) | .73 (.44) | .73 (.44) |
| Point of service (POS) | .09 (.28) | .08 (.27) |
| Preferred provider organization (PPO) | .07 (.25) | .07 (.26) |
| Medicaid | .03 (.17) | .03 (.17) |
| Exclusive provider organization (EPO) | .01 (.12) | .02 (.13) |
| Other insurance type | .07 (.26) | .07 (.26) |
| *Specialist Characteristics* | | |
| Female specialist | .26 (.44) | .24 (.43) |
| Specialist graduation year (mean) | 1988 (11) | 1988 (11) |
| Specialist experience with condition (mean) | .03 (.08) | .03 (.08) |
| Specialist top-15 medical school | .21 (.40) | .19 (.39) |
| Provider specialty | | |
| Dermatology | .18 (.39) | .17 (.38) |
| Orthopedic surgery | .16 (.37) | .11 (.31) |
| Otolaryngology | .06 (.24) | .06 (.24) |
| Urology | .07 (.26) | .05 (.23) |
| Surgery | .05 (.22) | .06 (.23) |
| Gastroenterology | .04 (.20) | .06 (.23) |
| Ophthalmology | .08 (.28) | .05 (.21) |
| Neurology | .05 (.21) | .04 (.21) |
| Cardiovascular disease | .05 (.23) | .04 (.21) |
| Obstetrics and gynecology | .05 (.22) | .03 (.17) |
| Other specialty | .19 (.39) | .32 (.47) |
| *Other Variables* | | |
| Female PCP | .44 (.50) | .43 (.50) |
| Proportion female specialists in HRR (mean) | .27 (.17) | .27 (.17) |
| Observations | 65,786 | 1,821,467 |

*Note:* This sample includes only index visits by adult patients who were referred by their PCP. The variables in this table are included as controls in subsequent regressions. Calendar year and diagnosis codes are omitted for brevity. The values in parentheses are the standard deviations from the mean. The SO versus non-SO differences in all variables in this table, except Medicaid coverage, are statistically significant based on *t*-tests and chi-square tests.

(McFadden 1973). Using a conditional logistic regression model allows us to limit the identifying variation to differences within each patient's choice set, which includes the specialists a patient can plausibly select. As a result, all patient (and PCP-level) attributes are conditioned out of the estimations. The regression model is specified as follows:

$$P\left(Y_{ik}=1|g_i,g_k,X\right)=\frac{e^{\beta 1_{g_i=g_k}+\delta X_{ik}}}{\sum_{k'\epsilon K_i} e^{\beta 1_{g_i=g_k}+\delta X_{ik}}} \quad (4)$$

where $Y_{ik} = 1$ if patient $i$ selects specialist $k$ and $Y_{ik} = 0$ otherwise. Specifically, selecting a specialist is modeled as a function of the patient and specialist having the same gender,

**Table 3**. Baseline Homophily First Visit

| | Probability of Specialist Selection | | | | |
|---|---|---|---|---|---|
| | 2010 | 2011 | 2012 | 2013 | 2014 |
| Female specialist | –.2831*** | –.2585*** | –.2373*** | –.2439*** | –.2292*** |
| | (.0041) | (.0049) | (.0052) | (.0052) | (.0060) |
| Same gender (patient) | .2292*** | .2254*** | .1989*** | .2057*** | .1921*** |
| | (.0042) | (.0051) | (.0054) | (.0054) | (.0063) |
| Same gender (PCP) | .1466*** | .1190*** | .1245*** | .1260*** | .1339*** |
| | (.0039) | (.0048) | (.0051) | (.0052) | (.0060) |
| Graduation year pre-1980 | .2286*** | .0550*** | –.0148** | –.0505*** | –.0993*** |
| | (.0052) | (.0058) | (.0061) | (.0061) | (.0070) |
| Graduation year 1980 to 2000 | .3831*** | .2257*** | .1664*** | .1290*** | .1001*** |
| | (.0047) | (.0050) | (.0051) | (.0051) | (.0055) |
| Distance (km) | –.0841*** | –.0838*** | –.0797*** | –.0771*** | –.0771*** |
| | (.0003) | (.0003) | (.0004) | (.0004) | (.0004) |
| *Case FEs* | Yes | Yes | Yes | Yes | Yes |
| *N* cases | 540,412 | 367,335 | 317,372 | 300,183 | 227,904 |
| *N* observations | 99,849,323 | 63,841,041 | 52,421,404 | 50,798,758 | 35,312,797 |
| Pseudo $R^2$ | .17 | .16 | .16 | .16 | .16 |

*Note:* All models are estimated using conditional logits.
*$p < .05$; **$p < .01$; ***$p < .001$ (two-tailed tests).

$1_{gi} = {}_{gk}$, and other specialist and patient-specialist relational attributes, $X_{ik}$.

The data consist of an observation for each patient *i*/PCP *j* – specialist *k* dyad (*i*/*j, 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 index visit generates a realized patient-specialist dyad, and for each realized referral, we create a set of counterfactual matches that consists of specialists the patient could have visited but did not. To establish the counterfactual matches, we match each chosen specialist with non-chosen specialists from the same hospital referral region, in the same specialty, and who practiced in the same month as the specialist to which patient *i* was referred by PCP *j*.

The main variable of interest in our model is *same gender (patient)*, which is a dummy variable that equals one if the patient and specialist are of the same gender. The conditional logistic regression we estimate is similar to a logistic regression with fixed effects, which means stable characteristics of the patient, such as gender and age, are differenced out (Chamberlain 1980). Therefore, the homophily effect in the regressions can only be identified as the average of male and female homophily: $\beta = (1/2)(\beta_{male} + \beta_{female})$. We also include a dummy for the gender of the specialist to capture the rate at which male specialists might be preferentially selected by both male and female patients. Other covariates in the regression include *same gender (PCP)*, set equal to one for PCP-specialist pairs of the same gender; *graduation year pre-1980*, set to one if the specialist graduated from medical school prior to 1980; *graduation year 1980 to 2000*, set to one if the specialist graduated from medical school between 1980 and 2000; and *distance (km)*, which is the distance in kilometers between the patient and the specialist.

Table 3 shows the results. Estimating the conditional logit is computationally intensive because of the large number of potential, non-occurring dyads. To keep each regression computationally feasible, we estimate the matching model by year rather than in pooled cross sections.[15] The results are similar across years. Based on the median effect

**Table 4**. Second Opinion Rate and Gender

| | Probability of Second Opinion after First Consult | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Female specialist | .003*** | .003*** | .004*** | .004*** | .007*** | .006*** |
| | (.000) | (.000) | (.000) | (.000) | (.001) | (.001) |
| Female patient | | | | .002*** | .003*** | .002*** |
| | | | | (.000) | (.000) | (.000) |
| Female specialist × Female patient | | | | | −.005*** | −.005*** |
| | | | | | (.001) | (.001) |
| Controls included | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| Specialty FEs | No | Yes | Yes | Yes | Yes | Yes |
| Diagnosis FEs | No | No | Yes | Yes | Yes | Yes |
| PCP FEs | No | No | No | No | No | Yes |
| *N* observations | 1,887,253 | 1,887,253 | 1,887,253 | 1,887,253 | 1,887,253 | 1,884,856 |

*Note:* All models are estimated using OLS. Logistic regressions yield similar results.
*$p < .05$; **$p < .01$; ***$p < .001$ (two-tailed tests).

size (column 4, year 2013), all else equal, the odds of a patient seeing a specialist of the same gender are 22 percent higher than the odds of seeing a specialist of the opposite gender. Note that this effect size is net of gender concordance between the PCP and specialist. PCPs also exhibit preferences for specialists of their gender, but the effect is less pronounced. The odds of a PCP referring to a specialist of the same gender are 13 percent higher (Model 4, year 2013) than the odds of referring to an opposite-gender specialist. Finally, physician experience and distance are also 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 summary, the baseline estimates show a strong, homophilous preference in specialist choice. 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 FOs stacks the deck against finding any gender differences in SOs. This is because patients who hold a strong preference for a same-gender specialist are likely to have selected a same-gender provider *in the first stage of the referral process*. When we observe male patient–female specialist dyads in first visits, we believe it is safe to assume the majority of male patients in these dyads have a weaker-than-average preference for a same-gender (male) provider. The bottom line is that our analysis likely provides a conservative estimate of the effect of patient-specialist gender matches on patient-initiated SOs.

## RESULTS

### *Main Analyses*

Table 4 reports six ordinary least squares (OLS) regression models in which the binary outcome captures whether patients secured an SO following an index visit.[16] We report heteroskedasticity robust standard errors. The regressions include all control variables listed in Table 1 and the fixed effects listed at the bottom of the table.

Column 1 in Table 4 includes only the gender of the specialist and shows patients of female FO specialists request SOs more

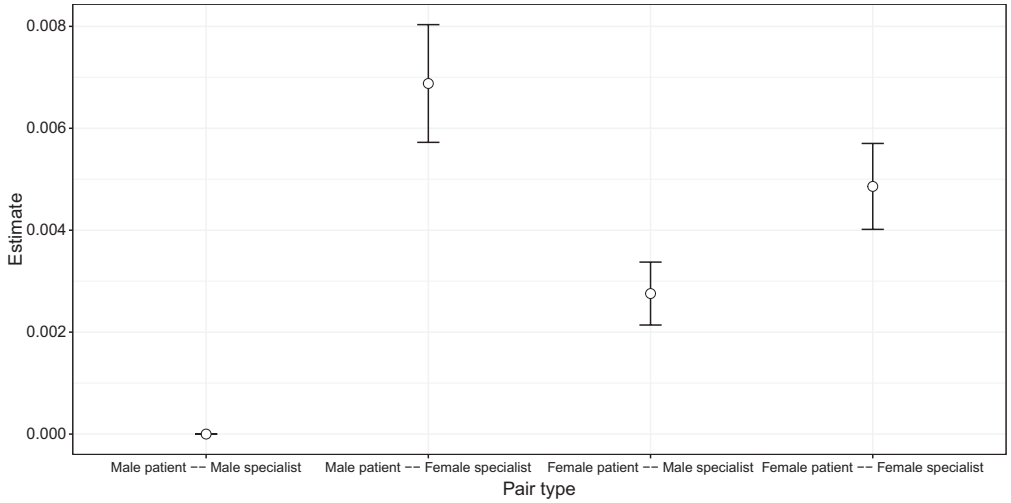


**Figure 2.** Estimated Effect Sizes – Extensive Margin
*Note:* This figure shows the change in SO rate by gender-pair type: female specialist–female patient, female specialist–male patient, male specialist–female patient, and male specialist–male patient. Male-male is the reference group. The graph shows that the behavior of male patients, compared to female patients, diverges much more depending on the gender of the FO specialist.

frequently than do patients of male specialists. In accord with Hypothesis 1, there is a difference in the rate of SOs as a function of specialist gender. The effect persists after including physician specialty FEs and diagnostic code FEs. Models 4 and 5 include the gender of the patient and an interaction between gender of the patient and gender of the FO specialist. The results show that female patients are more likely to seek SOs compared to male patients, but male patients' decision-making is much more sensitive to the gender of the specialist. This result supports Hypothesis 2. In Model 6, we incorporate PCP fixed effects, which do not alter the results.

Figure 2 illustrates the central pattern of results, showing the point estimates for each of the four possible gender pairings based on the coefficients in column 5, Table 4. The graph demonstrates 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 pairing is the reference group in the regressions and in the figure. It represents the dyadic gender composition with the lowest rate of SOs. As we anticipated in Hypothesis 2, the difference in the rate at which patients seek SOs after visiting a male versus female specialist is substantially higher for male patients than for female patients.

Turning to magnitudes, the effect sizes are meaningful. The coefficient estimates from column 3, Table 4, imply that patients seeing a female FO specialist are 11 percent more likely to seek an SO. Column 5, Table 4, shows that female patients visiting a female FO specialist are 6 percent more likely to seek an SO, and male patients visiting a female specialist are 20 percent more likely to seek an SO. The comparison group for these estimates is male patients seeing male specialists. To put these estimates into perspective, a one standard deviation increase in the Charlson score is associated with an 11 percent increase in the rate of SOs.

One concern is that despite the inclusion of a comprehensive set of control variables, the gender effects in Table 4 may be driven by unobserved differences in specialist quality or perceptions of risk. To better characterize the likelihood that unobserved confounders could drive the results, we evaluate how significant these omitted variables would need to be to wipe out the main effect of female specialist. To do this, we use Oster's (2019) methodology, which compares the coefficient

**Table 5**. Test for Importance of Unobserved Confounders

| | Probability of Second Opinion after First Consult | |
|---|---|---|
| | No Controls | With Controls |
| Female specialist | .00262 | .00364 |
| | (.00031) | (.00035) |
| $R^2$ | .000 | .009 |
| δ | | –19.3 |
| *N* observations | 1,887,253 | 1,887,253 |

*Note:* Estimates with versus without controls based on Model 5, Table 4. Column 1 reports estimates from regressing seeking an SO on the gender of the specialist without controls. Column 2 adds controls. The row for δ shows the amount of confounding from unobservables necessary, relative to the amount of confounding from control variables, to explain away the coefficient for female specialist. For the calculation of this δ, we use the Stata command *-psacalc-*. The researcher needs to determine what the practical maximum $R^2$ is. We set Rmax to 1.5 times the $R^2$ shown in the second column, which is a conservative value based on prior work by Oster (2019).
$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$ (two-tailed tests).

of interest and the $R^2$ between regressions with and without control variables. Specifically, we calculate $\delta$, which is the ratio of the influence of unobservables to the influence of included controls that would be required to drive the coefficient on gender (the treatment variable) to zero. Table 5 shows the results of this analysis for the probability of seeking an SO.

Oster (2019) suggests a $\delta > 1$ or $\delta < 0$ can be considered robust. Table 5 shows that the inclusion of controls moves the coefficient on female specialist *away from zero*.[17] This means that if unobservable confounders would behave similarly, our main estimate may represent a lower bound. Moreover, the corresponding $\delta$ implies that to explain away the impact in column 2, unobservables would have to move the coefficient in the opposite direction as observables, and their influence would have to be 19 times as large. Taken together, the results in Table 5 strongly suggest that omitted variable bias is unlikely to explain the results.

In a final step to rule out unobserved confounders, we estimate two regressions, one for male and one for female specialists, that include a full suite of physician fixed effects.[18] Table 6 shows the results. In interpreting the results, note that the baseline rate of seeking an SO is different in the two samples. The results suggest that if the FO specialist is female, female patients are 5 percent less likely than male patients to obtain an SO. If the specialist is male, male patients are 8 percent less likely than female patients to request an SO.[19] This is a large difference of over 60 percent (–.0789 / –.0489 = 1.6135). Mirroring the previous findings, when we include specialist fixed effects, we continue to find that male patients' decisions to seek an SO seem to be more strongly guided by the gender of the expert than are female patients' decisions. This occurs even when we limit variation to the gender of the patient and include fixed effects for all specialists.

In addition to our tests to evaluate whether confounding effects present a problem for our main results, we also evaluate whether sampling decisions influence the findings. In Appendix B, we re-estimate the coefficients from column 5, Table 4, using alternative definitions of SOs and removing observations for health conditions that are gender-specific, such as ovarian or prostate conditions. Across these and other robustness tests, the results are consistent with the findings in the main regressions. Finally, we take a slightly different approach to evaluate the concern that male and female specialists differ in quality. Specifically, we examine a health outcome: whether the patients of male and female

**Table 6**. Second Opinion Rate and Gender

| | Probability of Switching Specialist Gender | |
|---|---|---|
| | Female Specialists | Male Specialists |
| Female patient | −.002** | |
| | (.001) | |
| Male patient | | −.003*** |
| | | (.000) |
| Controls included | Yes | Yes |
| Year FEs | Yes | Yes |
| Specialty FEs | Yes | Yes |
| Diagnosis FEs | Yes | Yes |
| *N* observations | 459,323 | 1,423,873 |

*Note:* Both models are estimated using OLS. Regressions include a full set of specialist fixed effects. Column 1 includes all index consultations in which the FO clinician is female. Column 2 is for male FO specialists. Logistic regressions yield similar results.
*$p < .05$; **$p < .01$; ***$p < .001$ (two-tailed tests).

specialists have different 30-day hospital readmission rates. This is a standard measure of the quality of care in medical claims data. The results in Appendix B show no gender difference, which along with the physician fixed-effects regressions and the coefficient stability analysis, suggest that unobserved quality differences between male and female clinicians are very unlikely to drive the central results.

## SUPPLEMENTARY ANALYSES

We conduct two additional analyses that extend our findings and are important in their own right. First, we ask whether patients who seek SOs change the gender of the specialist they visit between the FO and the SO. Second, we estimate the effect of questioning expertise on the consumption of medical services, and therefore physician billings.

### *Specialist Gender Switching*

Do patients who seek SOs alter their choice of specialist gender between consultations? In other words, 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 7 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. This table includes all index visits that resulted in an SO. The table shows there is much less gender switching when the FO clinician is male compared to when the FO is female. Moreover, men are much more likely to switch specialist gender for the SO than are women, *if the specialist in the first visit was female*. Although female patients also are more likely to switch physician gender when the first specialist is female, the discrepancy is much less pronounced compared to male patients. Figure 3 depicts these differences. In short, the gender-switching results mirror the expectations we would have based on the theorized mechanisms.

### *Billings Implications*

The results presented so far establish that patients are more likely to exhibit behaviors consistent with a lack of confidence in the advice given by female specialists, and this result is driven mostly by male patients. Also, when patients seek SOs, they frequently switch from female to male providers. A question remains: what is the impact of these

**Table 7**. Gender Switching in Second Opinions

| | Probability of Switching Specialist Gender | |
|---|---|---|
| | (1) | (2) |
| Female specialist | .294*** | .420*** |
| | (.005) | (.008) |
| Female patient | | .060*** |
| | | (.004) |
| Female specialist × Female patient | | –.190*** |
| | | (.009) |
| Controls included | Yes | Yes |
| Year Fes | Yes | Yes |
| Specialty Fes | Yes | Yes |
| Diagnosis FEs | Yes | Yes |
| *N* observations | 65,786 | 65,786 |

*Note:* Both models are estimated using OLS. Logistic regressions yield similar results.
*$p < .05$; **$p < .01$; ***$p < .001$ (two-tailed tests).

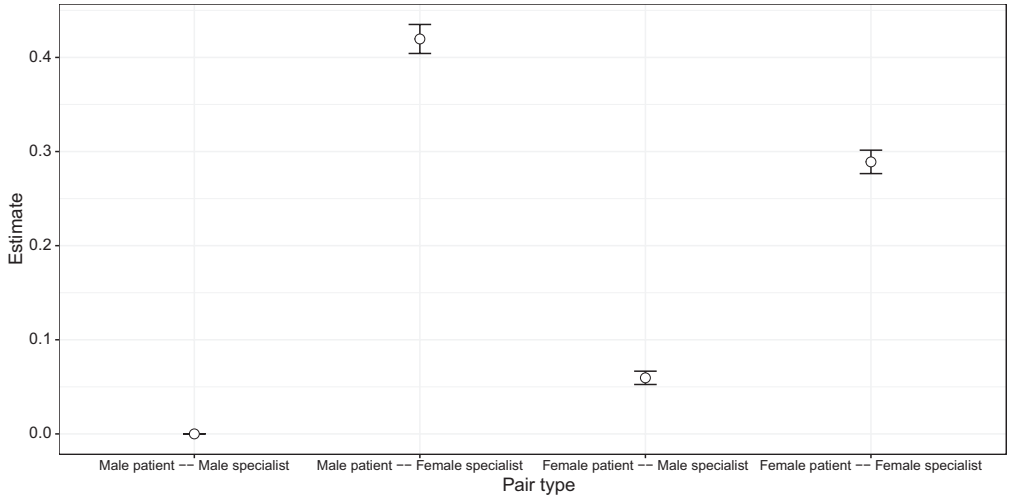


**Figure 3.** Estimated Effect Sizes – Intensive Margin
*Note:* This graph shows gender switching for the sample of SOs, for all gender pairings. Male patients seeing female specialists for a first consultation, are, by a wide margin, the most likely to switch to a male clinician for an SO.

dynamics on the medical spending by patients seen in the index visit on services provided in the following year by female versus male FO specialists?[20]

To answer this question, we rely on the fact that patients who lack confidence in the medical advice received are much less likely to return to the specialist who provided the recommendation than are patients who trust the advice. As shown in Table 2, patients who did not seek an SO spent almost 50 percent more on services provided by the FO specialist in the year following the index visit than did patients who sought an SO. As an upper-bound estimate of the impact of "questioning advice" on specialist billing, we estimate the difference in subsequent-year, aggregate billings for all patients who see male and female

**Table 8**. Differences in One-Year Billings

| | Log(one-year billing) | |
|---|---|---|
| | (1) | (2) |
| Female specialist | −.107*** | −.180*** |
| | (.008) | (.013) |
| Female patient | −.080*** | −.102*** |
| | (.007) | (.008) |
| Female specialist × Female patient | | .108*** |
| | | (.015) |
| Controls included | Yes | Yes |
| Year FEs | Yes | Yes |
| Specialty FEs | Yes | Yes |
| Diagnosis FEs | Yes | Yes |
| *N* observations | 1,887,253 | 1,887,253 |

*Note:* Both models are estimated using OLS, and the outcome variable is computed as the logarithm of the one-year billing amount.
*$p < .05$; **$p < .01$; ***$p < .001$ (two-tailed tests).

specialists following each FO consultation. It is important to note that even if we adjust this estimate for many controls, a significant billing difference between male and female specialists may be caused by factors in addition to patients' lack of confidence in advice given. For example, as a matter of practice style, male specialists may recommend more frequent follow-up visits compared to female specialists. Identifying persuasive controls for all such physician-specific differences is challenging. Therefore, the billing gap between male and female specialists we report may be best understood as an upper limit on the billing implications of a gender-based lack of confidence in physician expertise. However, we do note that unobserved specialist-specific differences in practice styles—in the absence of patients' gender-specific tendencies to question advice—should not lead to male and female patients utilizing different amounts of care as a function of specialist gender.

In Table 8, we again use the sample of 1.9 million index visits, but we now regress the log of the subsequent-year's total billing amount of each patient of the FO specialist on the gender of the patient, the gender of the specialist, and the interaction between the two. Model 1 shows that female specialists, on average, bill 10.7 percent less to their patients in the year following the FO visit than do their male counterparts. This difference is quite large but significantly smaller than the unadjusted mean difference in one-year patient billings reported in Table 1. The reduction in magnitude occurs because the regressions condition on physician gender-sorting into specialty, differences in patient health status by physician gender, and multiple other factors.

Model 2 allows us to evaluate whether the gender gap is the same for male versus female patients. It is not. When a female specialist sees a male patient, the average one-year billing amount is 18 percent less compared to a male specialist seeing a male patient, and this difference is only 7 percent for female patients. Note that in addition to providing a coarse estimate of how gender differences in trust in an expert affect the billings generated by specialists, these findings also perfectly replicate the findings presented in our main models but using patient-level billings as the outcome measure.

In summary, the billing results demonstrate significant financial implications of male and female patients varying in the degree to which they trust the advice of female

clinicians. Even if the estimates for the difference between male and female specialists are upwardly biased as a result of omitted variables, the substantial difference in effect sizes between male and female patients suggests that lack of full confidence in the advice of female experts likely contributes to the pay gap between male and female experts.

## DISCUSSION AND CONCLUSIONS

This article shows that patients differ in their responses to medical consultations with specialist providers. We find that the gender of the expert patients consult for a first opinion on a new health condition strongly influences the patient's decision to pursue an SO. Clients who consult female experts are more likely to obtain an SO, and this is particularly so if the client is male. Suggestive calculations highlight the potential consequences for physician remuneration of these gendered client preferences: if SOs shift the utilization of medical services from the first specialist to the less-likely-to-be female, second clinician, female specialists are less able to generate billings. Therefore, the incidence of "questioning expertise" in client-expert dyads likely accounts for a material fraction of the attainment gap among physicians—an occupation that continues to exhibit among the highest gender differences in earnings (Boulis and Jacobs 2011).

Scholars have identified many mechanisms that contribute to the gender pay gap in professions, including the presence of outright gender bias, childcare expectations and implied workforce attachment, and gender-based sorting into higher- versus lower-paying sub-specialties within occupations. All of these mechanisms have been repeatedly documented in professional labor markets, and most have been shown in medicine specifically. However, we believe our results suggest the side-by-side operation of more subtle processes related to the formation of confidence in advice given in client-expert dyads: possibly through multiple avenues, gendered patterns of interaction in these pairings influence the perceived value of expert services and thus clients' choice of service providers.

Our findings add nuance to prior work showing that, in allocative settings, both male and female evaluators tend to penalize women, particularly in contexts when merit is difficult to discern. In dyadic interactions where there is a high need for trust, gender concordance may engender a level of comfort that attenuates the general tendency for female providers to be penalized by both male and female clients. Although we are unable to fully disentangle effect magnitudes for status-based and similarity-based mechanisms, the fact patterns in our results suggest that trust facilitated by gender similarity affects how female experts are viewed. In short, both mechanisms appear to be operative in these data.

Assuming the patterns we see in the data do in fact indicate gender differences in the trust clients place in expert advice, what specific mechanisms might account for the results? The most straightforward possibility is an overt bias against female experts. Although this surely occurs, we believe it mostly unfolds at the first stage of selecting experts for consultations. As we show in Table 3, there is strong evidence of gender preferences in the selection of specialists among both male and female patients. We think it is safe to assume most patients who hold a strong predisposition to consult a gender-matched expert will exercise this preference in their selection of the FO specialist. As a result of this matching process, the patterns identified in our study are likely to be generated by patients who hold relatively weak preferences for male versus female providers. If strong preferences are already represented in the formation of the gender pairings in the FOs that constitute the risk set for SOs, what sources of variation remain?

One possibility that looms large is that gender dynamics affect what actually occurs during patient-physician consultations, such that the confidence patients have in the advice

given is influenced. Much prior research examines differences in practice and communication styles between female and male physicians. In this work, the preponderance of evidence suggests that in the daily practice of medicine, female and male physicians exhibit more similarities than differences. However, in the nuanced, communicative, and relational aspects of doctor-patient interactions, there is much more evidence of gendered dynamics. Summarizing the literature, Boulis and Jacobs (2011:13) conclude, "Gender matters far more with respect to time spent with patients and communication styles than it does with respect to diagnosis or treatment regimes."

Contextualizing these results within broader work on the sociology of the professions, much of the literature views expertise as an attribution that arises from professional credentialing (Collins and Evans 2008). This focus has led to work on the professions as a type of organizational form, with jurisdictional boundaries demarcated by interprofession jostling for the monopolization of esoteric knowledge (Abbott 1988). In consequence, some scholars argue that the sociology of the professions overemphasizes the analysis of organizational and institutional forms relative to the content and context of what experts actually do (Azocar and Ferree 2015). In accord with more microanalytic and symbolic interactionism approaches to the study of the professions, we believe the here-and-now context of exchanges in professional settings—the provision of expert advice in dyadic interactions that are rife with uncertainty—is a breeding ground for status and gender dynamics. In turn, these dynamics have strong implications for earnings attainment.

This study also illustrates how a sociological perspective can enrich our understanding of important health behaviors. In addition to the incidence of SOs, many other behaviors likely depend on the trust and confidence patients place in expert consultations. For example, medication adherence—consuming the medications prescribed by a physician—is a significant health behavior. We know that there is a very high incidence of patient nonadherence to prescribed medications, but few large-sample studies present cogent evidence of why some patients follow recommended drug regiments and others do not (Brown and Bussell 2011). Likewise, physical therapists, psychologists, and other healthcare professionals recommend many behaviors to their patients, but compliance with advice is variable. Advancing our knowledge about the formation of trust and confidence in client-expert dyads will likely inform the choices patients make about these and myriad other health behaviors.

In summary, our findings highlight how the perceived value of expert advice varies with the social fabric that characterizes client-expert interactions. We drew on two perspectives of how clients overcome the uncertainty endemic in the consumption of professional services, and we documented the financial implications for male versus female experts. Our data capture important social dynamics in the healthcare setting, but many other client-based professions share the boundary conditions that establish the limits of the theory, and we anticipate that the hypothesized relationships will generalize to these settings. In legal services, financial planning, auditing services, civil engineering, and many other client-based professions, social and interpersonal dynamics contour the perceived value of expertise. And like so many dimensions of life, perception has a way of becoming reality.

## APPENDIX

### *Part A: Validating the "Second Opinion" Label*

We perform several analyses to establish that the index visits resulting in SOs have characteristics that systematically distinguish them from other FO consultations. First, the existing, survey-based literature suggests that SOs should be more prevalent for serious health conditions. To assess this, we compute the average for three proxies of severity of a

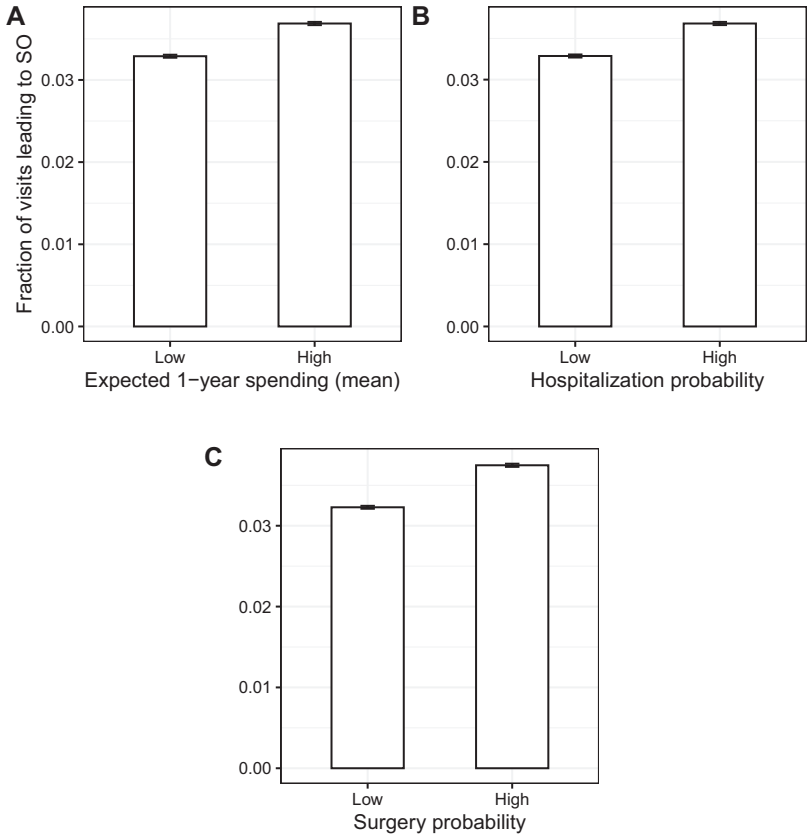


**Figure A1.** Second Opinions and the Severity of a Diagnosis
*Note:* Graphs show the fraction of index visits leading to an SO, broken out by one-year spending, hospitalization rate, and surgery rate associated with the main diagnostic code assigned to the patient for first opinion visit. Panel A assigns an index visit the mean one-year spending for the assigned diagnosis; panel B is the one-year probability of hospitalization for a diagnosis; and panel C is the one-year probability of surgery for a diagnosis. For expected spending, hospitalization, and probability of surgery, we use a median split of the sample to create "high" and "low" categories. Error bars represent the standard errors of the mean.

health condition: the one-year medical spending, the one-year hospitalization probability, and the one-year surgery probability associated with every *diagnosis code* in the claims data. The intuition is that patient-initiated SOs should be more common if the medical condition diagnosed in the first visit is costly to treat, is likely to lead to a hospitalization, 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 per-patient mean of this sum, for all diagnoses. For the hospitalization probability, we use the same strategy but instead record whether a patient is admitted to the hospital within a year after being first assigned diagnosis *i*. Finally, to compute the one-year surgery probability, we identified whether patients undergo a surgical procedure in the year following the initial diagnosis.[21] Figure A1 shows the results. As anticipated, if patients receive a diagnosis with high expected spending, a high likelihood of hospitalization, or a high likelihood of surgery, they are more likely to seek an SO consultation.

Another concern not necessarily addressed by this analysis is that SOs may occur because of a mismatch between the expertise of the first-seen physician and the expertise needed to successfully treat a patient's health condition. Medical fields can have elaborate sub-specializations, and for this reason or because a PCP's initial diagnosis turns out to be incorrect, patients occasionally may be mis-assigned to specialists in the referral process.[22] Note that in constructing the set of index visits, this is why we exclude all SOs that originate from a referral 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 assess the possibility that the SO cases among our index visits include a large fraction of expertise mismatches to patient conditions (and therefore represent misattributions of SOs to patient initiative), we extract treatment histories of specialists consulted in the FO visit to compare them to those of the specialist in the SO visit. The intuition for this analysis is that, if patients are frequently referred to specialists who do not have the expertise to treat their conditions, the treatment experience of the FO and SO specialists should be more dissimilar to one another than would be true for a benchmark of pairs of same-specialty physicians chosen at random. Conversely, if the FO-SO pair of physicians handle similar cases, and therefore represent suitable alternatives for first and second opinions for a given diagnosis, the treatment histories of the actual pair of FO-SO specialists should be considerably more similar than those of a randomly matched pair of providers.

We conduct this analysis by extracting vectors of frequencies of procedure codes performed by each pair of specialists and then computing the cosine similarity between the experience vectors of the FO and SO specialists.[23] To benchmark the resulting distribution of similarities, we also compute distances between the one-year treatment history of the FO specialist and two randomly sampled specialists. In one approach, we randomly select a specialist in the same specialty who also treated patients on the same day as the index visit. In a second approach, we match specialists on the diagnosis they most commonly treat and then randomly sample a specialist in the same specialty who also treated patients on the same day as the index visit.

Figure A2 shows the three distributions of similarity scores. Here, the results are remarkable. We find that, on average, the actual pairs of FO and SO specialists are much more similar in expertise than are pairs with counterfactually assigned specialists placed into the 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 that mismatches between specialist expertise and health condition in the index visit are unlikely to be common, and patient mis-assignment is not a central component of the data-generating process in the sample of SOs used here.

The final analysis we conduct to validate the sample of SOs considers the possibility that patients experience a new, related health condition between the first and second specialist visits. If this occurs frequently, it will conflate the sample of patient-initiated SOs. An individual who experiences a second health problem in the specialty of the first provider and then chooses another clinician to provide care is not seeking an SO for the original medical consultation. We therefore identify cases in which patients in the SO sample (1) visit their PCPs between consultations one and two, and (2) are assigned a primary diagnostic code that differs from the one assigned in the FO specialist visit. We find a PCP-initiated diagnosis code change in less than 10 percent of the cases in the SO sample.[24] The low percentage of cases in which there is a code change bolsters our confidence

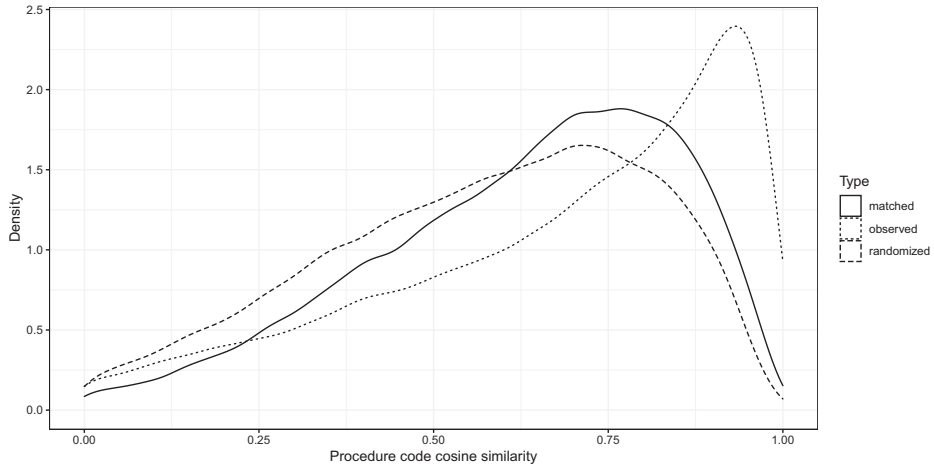


**Figure A2.** Distribution of Cosine Similarities between Observed, Matched, and Randomly Assigned Pairs of Specialists

*Note:* The dotted line shows the distribution of similarities between the treatment histories of the FO and SO specialist. As pairs of physicians become more similar, they trend toward the right side of the graph. The dashed and solid lines show the two counterfactual scenarios. Compared to the two counterfactual distributions, the observed distribution is shifted to the right and peaks near 1, suggesting much higher levels of expertise similarity in actually occurring, FO-SO physician pairs.

in the procedure for identifying SOs, but to err on the side of caution, we re-estimate our main regression model excluding these cases (see Appendix B, Table B2, Model 2).

## *Part B: 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 argued that a significant fraction of patient-initiated SOs occur because the client lacks full confidence in the guidance of the first expert, based on the specialist's gender. An alternative possibility is that the medical advice of the average female specialist is of lower quality than the advice of male specialists. Through the inclusion of a detailed set of controls, fixed-effects regressions, and our coefficient stability analysis, we demonstrated it is unlikely that such differences drive our results. However, to err on the side of caution, we conduct one additional analysis to evaluate this concern.

Specifically, in Table B1 we report the difference in 30-day hospital readmission rates of patients as a function of the gender of the FO specialist. Readmission rates are a common measure of health outcomes and have been targeted by policy interventions in an attempt to reduce healthcare costs (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 (for any reason) in the year following the index visit and were readmitted to the hospital within 30 days of the initial admission date. We limit these hospitalizations to those labeled "emergency" or "urgent" to ensure we are not picking up follow-up visits that could vary with the type of procedure recommended. Like in all other regression models, we condition on patient age and the Charlson comorbidity score, and we include provider-specialty fixed effects. Table B1 compares readmission rates between male and female specialists for the full sample of index visits. The regression shows a negative coefficient for female specialist, but the effect is not significantly different from zero. We therefore conclude that this evidence leans against the possibility that the gendered SO behavior of

**Table B1**. Readmission Rates by Specialist Gender

| | Probability of 30-Day Readmission |
|---|---|
| Female specialist | −.0002 |
| | (.0002) |
| Female patient | −.0021*** |
| | (.0002) |
| Controls included | Yes |
| Year FEs | Yes |
| Specialty FEs | Yes |
| Diagnosis FEs | Yes |
| *N* observations | 1,887,253 |

*Note:* We estimated this model using OLS. Logistic regression yields similar results.
*$p < .05$; **$p < .01$; ***$p < .001$ (two-tailed tests).

patients is driven by actual quality differences in the advice provided by female versus male specialists.

Next, we evaluate the robustness of the main effects by revisiting the design choices that define the sample. Specifically, in Table B2 we replicate results for the core regression analysis presented in Table 4, Model 5, but we exclude (1) observations we believe to be at highest likelihood of being miss-classified as SOs, (2) observations for health conditions that are specific to male or female patients, and (3) observations for which we have little or no historical data (i.e., observations from calendar year 2010). Specifically, Models 1 to 4 evaluate whether different choices in how to define SO cases change our results. In Model 1, we remove all cases for which the actual pair of FO and SO 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 (see results in Figure A2). In Model 2, we exclude cases where patients (1) visit their PCP between the first and second specialist consultation, and (2) are assigned a diagnosis that is different from the one assigned during the FO visit with the specialist. Some fraction of these cases could represent FOs for a new condition, from a second specialist. In Models 3 and 4, we revisit the length of time permitted to elapse until an SO from the six-month window in the core results. In Model 3, we constrain an SO to occur within three months after the first consultation. In Model 4, we further reduce the time window to one month after the original consultation. Effect sizes change slightly, but the central results are stable across all specifications.

In Model 5, we evaluate whether it is likely the results are driven not by men questioning the advice of female specialists, but by men feeling less comfortable discussing personal health issues with female specialists. Medical consultations are uniquely personal and often involve health concerns about which clients may feel embarrassment or exhibit other emotional responses. This raises the possibility that the pattern of results we observe is less indicative of questioning expertise, and instead has something to do with patients' discomfort in ongoing consultation or treatment with a physician of a given gender.

To explore this possibility, we limit the index sample of first visits to include only diagnoses that are approximately equally likely for male and female patients. In other words, we eliminate all single-sex-concentrated health problems. Model 5 in Table B2 limits the sample to include only first consultations in which the diagnosis assigned to the patient has a female-to-male patient gender ratio between .4 and .6. When we restrict the analysis to health conditions that affect both genders almost equally, we continue to find an identical pattern of results: male patients consulting female specialists are most likely

**Table B2.** Second Opinion Rate and Gender – Sensitivity Analyses

| | Probability of Second Opinion after First Consult | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Female specialist | .004*** | .005*** | .005*** | .002*** | .007*** | .006*** |
| | (.000) | (.001) | (.001) | (.000) | (.001) | (.001) |
| Female patient | .002*** | .001*** | .001*** | .000 | .003*** | .003*** |
| | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) |
| Female specialist | −.003*** | −.004*** | −.004*** | −.002*** | −.005*** | −.005*** |
| × Female patient | (.001) | (.001) | (.001) | (.000) | (.001) | (.001) |
| Controls included | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| Specialty FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| Diagnosis FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| *N* observations | 1,859,779 | 1,871,529 | 1,869,982 | 1,847,906 | 905,309 | 1,308,248 |

*Note:* We estimated all models using OLS. Logistic regressions yield similar results.
*$p < .05$; **$p < .01$; ***$p < .001$ (two-tailed tests).

to obtain an SO. Although archival medical claims data do not directly capture emotional responses to opposite-gender consultations, these results do not suggest that embarrassment or related emotions drive the main result.

Finally, we explore potential problems associated with left censoring. The Massachusetts APCD data begin on January 1, 2010. For individuals who have their first specialist visit in 2010, we do not have a full year of data to establish 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 excluded all cases in which the FO visit took place before January 1, 2011. Note that this removes a large number of cases from the sample, because the sampling restriction that the index visit is the first observed consultation to a given specialty front loads many of the cases in the sample. However, despite the large reduction in sample size, the effect sizes remain remarkably stable.

## Acknowledgments

The authors would like to thank Solene Delecourt, Aruna Ranganathan, Sameer Srivastava, Abhishek Nagaraj, Douglas Guilbeault, Hannah Neprash, Ziad Obermeyer, Susan Nguyen, and Elizabeth Fontana, as well as participants in research seminars at Columbia, MIT, Harvard, and UC-Berkeley. The authors would also like to thank the editors of the *American Sociological Review* as well as three anonymous reviewers for their helpful comments and suggestions.

## ORCID iD

Mathijs de Vaan https://orcid.org/0000-0003-1971-7290

## Data and Code

The authors would like to thank CHIA for providing access to the MA APCD. Code files for this paper can be found at https://osf.io/nzue2/.

## Notes

1. Because patients may seek guidance from their PCP to decide whether to seek an SO, and because PCPs may vary in how much they rely on specialist gender in providing this guidance, we adjust the estimates for PCP-specific fixed effects.
2. Patients routinely choose not to adopt medical recommendations. Prior work finds that patients frequently opt not to take prescribed medication (Fischer et al. 2010; Koenig 2011), and many choose not to return for follow-up visits after a hospitalization (Gavish et al. 2015).
3. Note that we make no assumption about whether SOs improve or detract from health outcomes, or about the optimal base rate for SO consultations. Some of the literature indicates SOs improve patient health outcomes because treatment decisions are "wiser" (Barnett et al. 2019). Conversely,

other studies emphasize the cost of SOs that do not result in changes in treatment. We remain agnostic to these issues and certainly do not intend to imply that our results have implications for the optimal incidence of SO. Rather, we analyze this outcome because in a carefully constructed sample, we believe it to be indicative of patient confidence in advice received in clinical consultations.

4. There are partial exceptions to this. For example, patients with chronic medical conditions often conduct extensive research to understand their situations, and their lived experiences provide them with additional knowledge on their health condition (Prior 2003; Timmermans 2020). In this article, we study the behavior of patients following a first-time consultation with a physician in a given medical specialty that is novel to the focal patient, given that individual's health history. Therefore, it is unlikely the patients in our sample have developed deep expertise about their medical conditions.
5. We leverage the referral indicator in our data, which is used in both HMO and POS insurance plans as well as a few PPO plans. Each referral also lists the PCP (i.e., physicians with family medicine or internal medicine as their listed specialty) who provided the referral. Office visits are identified using Current Procedural Terminology codes (CPT) 992**.
6. The primary decision-maker in these consultations, a parent, may be influenced by specialist gender in the decision to pursue an SO. However, we do not observe the gender of the parent who is most influential in this decision, which prevents us from using these cases to examine the interaction of client and expert gender.
7. If anything, the results are slightly stronger if we reduce the size of the window (for details, see Appendix Table B2).
8. Because we exclude individuals with chronic medical conditions who consult a large number of specialists on an ongoing basis, this number understates the incidence of SOs in the overall medical claims data.
9. The diagnoses codes in our data are ICD-9 codes. ICD-9 is a hierarchical classification system with a maximum of five digits used by physicians to codify the medical conditions of their patients. We will discuss ICD-9 codes in more detail.
10. Note that our sampling strategy to only include a patient's first-observed visit to a specialty effectively suppresses the severity and complexity of underlying health conditions in the data. Chronically ill patients, on average, have the most severe and complex health issues, and most of their visits to a specialist are not first-time visits.
11. It is not possible to pool male and female specialists in the FE regressions because specialist gender will be collinear with the specialist fixed effects.
12. In an additional robustness check, we also compare male and female specialists by examining whether the health outcomes (as measured by 30-day readmissions) of their patients vary. The rationale for this analysis is that if the medical recommendations of male versus female specialists are substantially better, one should be able to demonstrate this in the data.
13. Sahni and colleagues (2016) show that specialization, rather than absolute experience, has a greater impact on quality of care.
14. Harvard Medical School, Johns Hopkins University School of Medicine, University of Pennsylvania School of Medicine, New York University School of Medicine, Columbia University College of Physicians and Surgeons, Stanford University School of Medicine, Mayo Medical School, University of California-UCLA School of Medicine, University of California-San Francisco School of Medicine, Washington University School of Medicine, Cornell University Medical College, Duke University School of Medicine, University of Washington School of Medicine, University of Pittsburgh School of Medicine, and University of Michigan Medical School.
15. The sample size declines each year because we restrict the data to include only patients who visit a specialist in a specialty for the first time within the data.
16. We report linear probability models for ease of interpretation, but logistic regressions yield similar results.
17. Note that the pattern we speculated about earlier (i.e., less healthy patients are more likely to select male specialists), may provide an explanation for this result.
18. The statistical significance of the estimates obtained from these two regressions cannot be directly evaluated because they are estimated from two different samples. Therefore, rather than looking for statistical similarity, our main goal for this analysis is to evaluate whether the results are broadly in line with the results presented in Table 4.
19. These percentages are calculated as follows: the SO base rate for female specialists is .0368. Therefore, female patients visiting a female specialist are 4.9 percent (–.0018 / .0368 = –.0489) less likely to seek an SO. The SO base rate for male specialists is .0342. Therefore, male patients visiting a female specialist are 7.9 percent (–.0027 / .0342 = –.0789) less likely to seek an SO.
20. We observe virtually all medical billings in Massachusetts but not physician income. Although income is not perfectly correlated with billing, prior work shows billing amounts have a large effect on physician earnings (Ryan et al. 2015).
21. We used the narrow definition developed by the Healthcare Cost and Utilization Project (HCUP) to identify CPT codes that indicate surgical interventions (https://www.hcup-us.ahrq.gov/toolssoftware/surgeryflags_svcproc/surgeryflagssvc_proc.jsp).

22. Orthopedists, for example, may specialize in a variety of sub-specialties, including joint replacement, spine surgery, and hand surgery.
23. Each medical claim for professional services (i.e., services performed by a medical provider) lists at least one procedure code. The listed procedure code forms the basis for reimbursement and therefore represents a high accuracy field in medical claims data.
24. Inspection of the data indicates that many of these cases occur because the PCP inputs a diagnostic code for the same health issue, but it is less specific than the one assigned by the specialist.

## References


Abascal, Maria, and Delia Baldassarri. 2015. "Love Thy Neighbor? Ethnoracial Diversity and Trust Reexamined." *American Journal of Sociology* 121(3):722–82.

Abbott, Andrew. 1988. *The System of Professions: An Essay on the Division of Expert Labor*. Chicago: University of Chicago Press.

Abrams, Dominic, and Michael A. Hogg. 1990. "Social Identification, Self-Categorization and Social Influence." *European Review of Social Psychology* 1(1):195–228.

Adams, Tracey L. 2010. "Gender and Feminization in Health Care Professions." *Sociology Compass* 4(7):454–65.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113(1):151–84.

Azmat, Ghazala, and Rosa Ferrer. 2017. "Gender Gaps in Performance: Evidence from Young Lawyers." *Journal of Political Economy* 125(5):1306–55.

Azocar, Maria J., and Myra Marx Ferree. 2015. "Gendered Expertise." *Gender & Society* 29(6):841–62.

Azoulay, Pierre, Toby Stuart, and Yanbo Wang. 2014. "Matthew: Effect or Fable?" *Management Science* 60(1):92–109.

Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva. 2017. "Does the Gender Composition of Scientific Committees Matter?" *American Economic Review* 107(4):1207–38.

Barley, Stephen R. 1989. "Careers, Identities, and Institutions: Legacy of the Chicago School of Sociology." Pp. 41–65 in *Handbook of Career Theory*, edited by M. Arthur, D. Hall, and B. Lawrence. Cambridge, UK: Cambridge University Press.

Barnett, Michael L., Dhruv Boddupalli, Shantanu Nundy, and David W. Bates. 2019. "Comparative Accuracy of Diagnosis by Collective Intelligence of Multiple Physicians vs. Individual Physicians." *JAMA Network Open* 2(3):e190096.

Becker, Marshall H. 1970. "Sociometric Location and Innovativeness: Reformulation and Extension of the Diffusion Model." *American Sociological Review* 35(2):267–82.

Beckman, Christine M., and Damon J. Phillips. 2005. "Interorganizational Determinants of Promotion: Client Leadership and the Attainment of Women Attorneys." *American Sociological Review* 70 (4):678–701.

Berger, Joseph, Bernard P. Cohen, and Morris Zelditch Jr. 1972. "Status Characteristics and Social Interaction." *American Sociological Review* 37(3):241–55.

Bian, Yanjie. 1997. "Bringing Strong Ties Back In: Indirect Ties, Network Bridges, and Job Searches in China." *American Sociological Review* 62(3):366–85.

Blau, Francine D., and Lawrence M. Kahn. 2013. "Female Labor Supply: Why Is the United States Falling Behind?" *American Economic Review* 103(3):251–56.

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

Bloom, Gerald, Hilary Standing, and Robert Lloyd. 2008. "Markets, Information Asymmetry and Health Care: Towards New Social Contracts." *Social Science & Medicine* 66(10):2076–87.

Boulis, Ann K., and Jerry A. Jacobs. 2011. *The Changing Face of Medicine: Women Doctors and the Evolution of Health Care in America*. Ithaca, NY: Cornell University Press.

Brewer, Marilynn B. 1979. "In-group Bias in the Minimal Intergroup Situation: A Cognitive Motivational Analysis." *Psychological Bulletin* 86(2):307–24.

Brewer, Marilynn B. 2008. "Depersonalized Trust and Ingroup Cooperation." Pp. 215–32 in *Rationality and Social Responsibility: Essays in Honor of Robyn Mason Dawes*, edited by J. I. Krueger. New York: Psychology Press.

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(12):4427–31.

Brown, Marie T., and Jennifer K. Bussell. 2011. "Medication Adherence: WHO Cares?" *Mayo Clinic Proceedings* 86(4):304–14.

Castilla, Emilio J. 2008. "Gender, Race, and Meritocracy in Organizational Careers." *American Journal of Sociology* 113(6):1479–526.

Chamberlain, Gary. 1980. "Analysis of Covariance with Qualitative Data." *The Review of Economic Studies* 47(1):225–38.

Chiou, Jyh-Shen, and Cornelia Droge. 2006. "Service Quality, Trust, Specific Asset Investment, and Expertise: Direct and Indirect Effects in a Satisfaction-Loyalty Framework." *Journal of the Academy of Marketing Science* 34(4):613–27.

Chiu, Chi-yue, Michael W. Morris, Ying-yi Hong, and Tanya Menon. 2000. "Motivated Cultural Cognition: The Impact of Implicit Cultural Theories on Dispositional Attribution Varies as a Function of Need for Closure." *Journal of Personality and Social Psychology* 78(2):247–59.

Collins, Harry, and Robert Evans. 2008. *Rethinking Expertise*. Chicago: University of Chicago Press.

Cook, Karen S., ed. 2001. *Trust in Society*. New York: Russell Sage Foundation.

Cook, Karen S., and Richard M. Emerson. 1978. "Power, Equity and Commitment in Exchange Networks." *American Sociological Review* 43(5):721–39.

Cook, Karen S., Roderick M. Kramer, David H. Thom, Irena Stepanikova, Stefanie Bailey Mollborn, and Robin M. Cooper. 2004. "Trust and Distrust in Patient-Physician Relationships: Perceived Determinants of High- and Low-Trust Relationships in Managed-Care Settings." Pp. 65–98 in *Trust and Distrust in Organizations: Dilemmas and Approaches*, edited by R. M. Kramer and K. S. Cook. New York: Russell Sage Foundation.

Correll, Shelley J., Cecilia L. Ridgeway, Ezra W. Zuckerman, Sharon Jank, Sara Jordan-Bloch, and Sandra Nakagawa. 2017. "It's the Conventional Thought That Counts: How Third-Order Inference Produces Status Advantage." *American Sociological Review* 82(2):297–327.

Dahlander, Linus, and Daniel A. McFarland. 2013. "Ties That Last: Tie Formation and Persistence in Research Collaborations over Time." *Administrative Science Quarterly* 58:69–110.

de Vaan, Mathijs, and Toby Stuart. 2019. "Does Intrahousehold Contagion Cause an Increase in Prescription Opioid Use?" *American Sociological Review* 84(4):577–608.

Del Rossi, Alison F., and Joni Hersch. 2020. "Gender and the Consulting Academic Economist." *Economic Inquiry* 58(3):1200–16.

Eddy, David M. 1984. "Variations in Physician Practice: The Role of Uncertainty." *Health Affairs* 3(2):74–89.

Ericson, Keith Marzilli, and Amanda Starc. 2015. "Measuring Consumer Valuation of Limited Provider Networks." *American Economic Review* 105(5):115–19.

Evans, Anthony M., and Joachim I. Krueger. 2009. "The Psychology (and Economics) of Trust." *Social and Personality Psychology Compass* 3(6):1003–17.

Fang, Margaret C., Ellen P. McCarthy, and Daniel E. Singer. 2004. "Are Patients More Likely to See Physicians of the Same Sex? Recent National Trends in Primary Care Medicine." *The American Journal of Medicine* 117(8):575–81.

Fischer, Michael A., Margaret R. Stedman, Joyce Lii, Christine Vogeli, William H. Shrank, M. Alan Brookhart, and Joel S. Weissman. 2010. "Primary Medication Non-adherence: Analysis of 195,930 Electronic Prescriptions." *Journal of General Internal Medicine* 25(4):284–90.

Flemons, W. Ward, Neil J. Douglas, Samuel T. Kuna, Daniel O. Rodenstein, and John Wheatley. 2004. "Access to Diagnosis and Treatment of Patients with Suspected Sleep Apnea." *American Journal of Respiratory and Critical Care Medicine* 169(6):668–72.

Foddy, Margaret, Michael J. Platow, and Toshio Yamagishi. 2009. "Group-Based Trust in Strangers: The Role of Stereotypes and Expectations." *Psychological Science* 20(4):419–22.

Fox, Renee C. 1957. "Training for Uncertainty." Pp. 207–41 in *The Student-Physician*, edited by R. K. Merton, G. Reader, and P. L. Kendall. Cambridge, MA: Harvard University Press.

Freidson, Eliot. 1988. *Profession of Medicine: A Study of the Sociology of Applied Knowledge*. Chicago: 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):1–12.

Gavish, Rachel, Amalia Levy, Or Kalchiem Dekel, Erez Karp, and Nimrod Maimon. 2015. "The Association between Hospital Readmission and Pulmonologist Follow-up Visits in Patients with COPD." *Chest* 148(2):375–81.

Goldin, Claudia, and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians." *American Economic Review* 90(4):715–41.

Goodman-Delahunty, Jane, Pär Anders Granhag, Maria Hartwig, and Elizabeth F. Loftus. 2010. "Insightful or Wishful: Lawyers' Ability to Predict Case Outcomes." *Psychology, Public Policy, and Law* 16(2):133–57.

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(7):1202–12.

Haines, Elizabeth L., Kay Deaux, and Nicole Lofaro. 2016. "The Times They Are A-changing . . . Or Are They Not? A Comparison of Gender Stereotypes, 1983–2014." *Psychology of Women Quarterly* 40(3):353–63.

Hardin, Russell. 2001. *Conceptions and Explanations of Trust*. New York: Russell Sage Foundation.

Heritage, John, and Douglas W. Maynard. 2006. "Problems and Prospects in the Study of Physician-Patient Interaction: 30 Years of Research." *Annual Review of Sociology* 32:351–74.

Howieson, Jill, and Shane L. Rogers. 2019. "Rethinking the Lawyer-Client Interview: Taking a Relational Approach." *Psychiatry, Psychology and Law* 26(4):659–68.

Hughes, Everett C. 1963. "Professions." *Daedalus* 92:655–68.

Johfre, Sasha Shen, and Jeremy Freese. 2021. "Reconsidering the Reference Category." *Sociological Methodology* 51(2):253–69.

Koenig, Christopher J. 2011. "Patient Resistance as Agency in Treatment Decisions." *Social Science & Medicine* 72(7):1105–14.

Kollock, Peter. 1994. "The Emergence of Exchange Structures: An Experimental Study of Uncertainty, Commitment, and Trust." *American Journal of Sociology* 100(2):313–45.

Koven, Suzanne. 2017. "Letter to a Young Female Physician." *New England Journal of Medicine* 376(20):1907–09.

Landon, Bruce E., and Robert E. Mechanic. 2017. "The Paradox of Coding: Policy Concerns in the Move to Risk-Based Provider Contracts." *The New England Journal of Medicine* 377(13):1211–13.

Light, Donald W. 2000. "The Medical Profession and Organizational Change: From Professional Dominance to Countervailing Power." *Handbook of Medical Sociology* 5:201–16.

Lundberg, Ian, Rebecca Johnson, and Brandon M. Stewart. 2021. "What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory." *American Sociological Review* 86(3):532–65.

Lurie, Nicole, Karen Margolis, Paul G. McGovern, and Pamela Mink. 1998. "Physician Self-report of Comfort and Skill in Providing Preventive Care to Patients of the Opposite Sex." *Archives of Family Medicine* 7(2):134–37.

Mann, Allison, and Thomas A. DiPrete. 2013. "Trends in Gender Segregation in the Choice of Science and Engineering Majors." *Social Science Research* 42(6):1519–41.

McFadden, Daniel. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." Pp. 105–42 in *Frontiers in Econometrics*, edited by P. Zarembka. New York: Academic Press.

Meeuwesen, Ludwien, Cas Schaap, and Cees Van Der Staak. 1991. "Verbal Analysis of Doctor-Patient Communication." *Social Science & Medicine* 32(10): 1143–50.

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(6):1013–36.

Melamed, David, Matthew Sweitzer, Brent Simpson, Jered Z. Abernathy, Ashley Harrell, and Christopher W. Munn. 2020. "Homophily and Segregation in Cooperative Networks." *American Journal of Sociology* 125(4):1084–1127.

Miller, Binny. 1994. "Give Them Back Their Lives: Recognizing Client Narrative in Case Theory." *Michigan Law Review* 93:485.

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(41):16474–79.

Neimanis, Ieva, Kathryn Gaebel, Robert Dickson, Richard Levy, Cindy Goebel, Angelo Zizzo, Anne Woods, and John Corsini. 2017. "Referral Processes and Wait Times in Primary Care." *Canadian Family Physician* 63(8):619–24.

Oakes, Penelope J., S. Alexander Haslam, and John C. Turner. 1994. *Stereotyping and Social Reality*. Hoboken, NJ: Blackwell Publishing.

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(1):36–43.

Olayiwola, J. Nwando, Daren Anderson, Nicole Jepeal, Robert Aseltine, Christopher Pickett, Jun Yan, and Ianita Zlateva. 2016. "Electronic Consultations to Improve the Primary Care–Specialty Care Interface for Cardiology in the Medically Underserved: A Cluster-Randomized Controlled Trial." *The Annals of Family Medicine* 14(2):133–40.

Oldmeadow, Julian A., Michael J. Platow, Margaret Foddy, and Donna Anderson. 2003. "Self Categorization, Status, and Social Influence." *Social Psychology Quarterly* 66(2):138–52.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37(2):187–204.

Parsons, Talcott. 1951. "Illness and the Role of the Physician: A Sociological Perspective." *American Journal of Orthopsychiatry* 21(3):452–60.

Payne, Velma L., Hardeep Singh, Ashley N. D. Meyer, Lewis Levy, David Harrison, and Mark L. Graber. 2014. "Patient-Initiated Second Opinions: Systematic Review of Characteristics and Impact on Diagnosis, Treatment, and Satisfaction." *Mayo Clinic Proceedings* 89(5):687–96.

Pedulla, David S., and Sarah Thébaud. 2015. "Can We Finish the Revolution? Gender, Work-Family Ideals, and Institutional Constraint." *American Sociological Review* 80(1):116–39.

Petersen, Trond, Ishak Saporta, and Marc-David L. Seidel. 2000. "Offering a Job: Meritocracy and Social Networks." *American Journal of Sociology* 106(3):763–816.

Pilnick, Alison, and Robert Dingwall. 2011. "On the Remarkable Persistence of Asymmetry in Doctor/Patient Interaction: A Critical Review." *Social Science & Medicine* 72(8):1374–82.

Pilnick, Alison, Jon Hindmarsh, and Virginia Teas Gill. 2009. "Beyond 'Doctor and Patient': Developments in the Study of Healthcare Interactions." *Sociology of Health & Illness* 31(6):787–802.

Podolny, Joel M. 2010. *Status Signals: A Sociological Study of Market Competition*. Princeton, NJ: 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(4):361–64.

Prior, Lindsay. 2003. "Belief, Knowledge and Expertise: The Emergence of the Lay Expert in Medical Sociology." *Sociology of Health & Illness* 25(3):41–57.

Prisbell, Marshall, and Janis F. Andersen. 1980. "The Importance of Perceived Homophily, Level of Uncertainty, Feeling Good, Safety, and Self-Disclosure in Interpersonal Relationships." *Communication Quarterly* 28(3):22–33.

Quadlin, Natasha. 2018. "The Mark of a Woman's Record: Gender and Academic Performance in Hiring." *American Sociological Review* 83(2):331–60.

Ridgeway, Cecilia L., and Shelley J. Correll. 2004. "Unpacking the Gender System: A Theoretical Perspective on Gender Beliefs and Social Relations." *Gender & Society* 18(4):510–31.

Ridgeway, Cecilia L., and Kristan Glasgow Erickson. 2000. "Creating and Spreading Status Beliefs." *American Journal of Sociology* 106(3):579–615.

Ridgeway, Cecilia L., and Lynn Smith-Lovin. 1999. "The Gender System and Interaction." *Annual Review of Sociology* 25(1):191–216.

Rivera, Lauren A. 2012. "Hiring as Cultural Matching: The Case of Elite Professional Service Firms." *American Sociological Review* 77(6):999–1022.

Rivera, Lauren A. 2015. *Pedigree*. Princeton, NJ: Princeton University Press.

Rudman, Laurie A., and Julie E. Phelan. 2008. "Backlash Effects for Disconfirming Gender Stereotypes in Organizations." *Research in Organizational Behavior* 28:61–79.

Ryan, Andrew M., Stephen M. Shortell, Patricia P. Ramsay, and Lawrence P. Casalino. 2015. "Salary and Quality Compensation for Physician Practices Participating in Accountable Care Organizations." *The Annals of Family Medicine* 13(4):321–24.

Sahni, Nikhil R., Maurice Dalton, David M. Cutler, John D. Birkmeyer, and Amitabh Chandra. 2016. "Surgeon Specialization and Operative Mortality in United States: Retrospective Analysis." *BMJ* 354:i3571.

Sandefur, Rebecca L. 2015. "Elements of Professional Expertise: Understanding Relational and Substantive Expertise through Lawyers' Impact." *American Sociological Review* 80(5):909–33.

Sandhu, Harbinder, Ann Adams, Laura Singleton, David Clark-Carter, and Jane Kidd. 2009. "The Impact of Gender Dyads on Doctor–Patient Communication: A Systematic Review." *Patient Education and Counseling* 76(3):348–55.

Sarsons, Heather. 2017. "Interpreting Signals in the Labor Market: Evidence from Medical Referrals." Working paper, University of Chicago, Booth School of Business.

Sharma, Anurag. 1997. "Professional as Agent: Knowledge Asymmetry in Agency Exchange." *Academy of Management Review* 22(3):758–98.

Shmueli, Liora, Nadav Davidovitch, Joseph S. Pliskin, Igal Hekselman, Ran D. Balicer, and Geva Greenfield. 2019. "Reasons, Perceived Outcomes and Characteristics of Second Opinion Seekers: Are There Differences in Private vs. Public Settings?" *BMC Health Services Research* 19:238.

Shmueli, Liora, Erez Shmueli, Joseph S. Pliskin, Ran D. Balicer, Nadav Davidovitch, Igal Hekselman, and Geva Greenfield. 2019. "Second Opinion Utilization by Healthcare Insurance Type in a Mixed Private-Public Healthcare System: A Population-Based Study." *BMJ Open* 9:e025673.

Sniezek, Janet A., and Lyn M. Van Swol. 2001. "Trust, Confidence, and Expertise in a Judge-Advisor System." *Organizational Behavior and Human Decision Processes* 84(2):288–307.

Song, Zirui, Thomas D. Sequist, and Michael L. Barnett. 2014. "Patient Referrals: A Linchpin for Increasing the Value of Care." *JAMA* 312(6):597–98.

Stuart, Toby E., Ha Hoang, and Ralph C. Hybels. 1999. "Interorganizational Endorsements and the Performance of Entrepreneurial Ventures." *Administrative Science Quarterly* 44(2):315–49.

Tajfel, Henri, Michael G. Billig, Robert P. Bundy, and Claude Flament. 1971. "Social Categorization and Intergroup Behaviour." *European Journal of Social Psychology* 1(2):149–78.

Tajfel, Henri, John C. Turner, William G. Austin, and Stephen Worchel. 1979. "An Integrative Theory of Intergroup Conflict." Pp. 33–37 in *The Social Psychology of Intergroup Relations*, edited by W. G. Austin and S. Worchel. Monterey, CA: Brooks/Cole.

Terry, Deborah J., and Michael A. Hogg. 1996. "Group Norms and the Attitude-Behavior Relationship: A Role for Group Identification." *Personality and Social Psychology Bulletin* 22(8):776–93.

Timmermans, Stefan. 2020. "The Engaged Patient: The Relevance of Patient–Physician Communication for Twenty-First-Century Health." *Journal of Health and Social Behavior* 61(3):259–73.

Webster, Murray, and James E. Driskell. 1983. "Beauty as Status." *American Journal of Sociology* 89(1):140–65.

West, Candace. 1984. "When the Doctor Is a 'Lady': Power, Status and Gender in Physician-Patient Encounters." *Symbolic Interaction* 7(1):87–106.

Williams, Melissa J., and Larissa Z. Tiedens. 2016. "The Subtle Suspension of Backlash: A Meta-analysis of Penalties for Women's Implicit and Explicit Dominance Behavior." *Psychological Bulletin* 142(2):165–97.

Zeltzer, Dan. 2020. "Gender Homophily in Referral Networks: Consequences for the Medicare Physician Earnings Gap." *American Economic Journal: Applied Economics* 12(2):169–97.

**Mathijs de Vaan** is an Assistant Professor at the Haas School of Business, University of California-Berkeley. His research investigates the causes and consequences of inequality in resources among scientists and the intersection of physician referral networks and variation in costs and quality of health care.

**Toby Stuart** is the Helzel Professor at the Haas School of Business, University of California-Berkeley. His recent research projects investigate social networks and entrepreneurship in population-level career histories, corporate fraudulent conduct in emerging markets, and social and physician referral networks and health outcomes.