

Gender Differences in Medical Evaluations: Evidence from Randomly Assigned Doctors ^{*}

Marika Cabral and Marcus Dillender [†]

May 2022

Abstract

While a growing body of evidence documents large gender disparities in health care and related social insurance programs, little is known about what drives these disparities. We leverage administrative data and random assignment of doctors to patients in workers' compensation insurance to study the impact of gender match between doctors and patients on medical evaluations and subsequent social insurance benefits received. Compared to differences among their male patient counterparts, female patients randomly assigned a female doctor rather than a male doctor are 5.0% more likely to be evaluated as disabled and receive 8.5% more subsequent cash benefits on average. There is no analogous gender-match effect for male patients. The magnitude of these effects implies that having female doctors evaluate patients entirely offsets the observed gender gap in the likelihood of being evaluated as disabled when male doctors evaluate patients. We explore mechanisms through further analysis of the administrative data and complementary survey evidence. In addition, we present broader evidence on gender gaps in workers' compensation insurance and gender homophily in patients' selections of doctors in settings where patients have choice. Combining this evidence, we conduct policy counterfactuals illustrating how policies increasing gender diversity among doctors or increasing gender homophily in patient-doctor matches may impact gender gaps in evaluated disability. Our findings indicate that policies increasing the share of female patients evaluated by female doctors may substantially shrink gender gaps in medical evaluations and associated outcomes.

^{*}For providing helpful comments, we thank Marcella Alsan, Sandra Black, Amitabh Chandra, Seema Jayachandran, Adriana Lleras-Muney, Heidi Williams, as well as participants of the NBER Summer Institute Health Care Meetings 2021, the Chicago Booth Junior Health Economics Summit 2020, and the ASSA annual meetings 2021. We thank Daniel Jordan Alvarez, Bokyoung Kim, and Jinyeong Son for their excellent research assistance. Cabral gratefully acknowledges financial support from the National Science Foundation CAREER Award (1845190). Cabral and Dillender gratefully acknowledge financial support for this research from the US Social Security Administration. The research reported herein was performed pursuant to grant RDR18000003 from the US Social Security Administration (SSA) funded as part of the Retirement and Disability Research Consortium. The opinions and conclusions expressed are solely those of the author(s) and do not represent the opinions or policy of SSA, any agency of the Federal Government, or NBER. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the United States Government or any agency thereof.

[†]Cabral: University of Texas at Austin Department of Economics and NBER, marika.cabral@utexas.edu. Dillender: University of Illinois at Chicago Division of Health Policy and Administration School of Public Health and NBER, modillen@uic.edu.

1 Introduction

Gender disparities are observed in many settings ranging from education to labor markets to financial transactions. A growing body of evidence from medical studies suggests these disparities extend to health care settings, where recent work has shown that female patients—relative to male patients—receive less health care for similar medical conditions and are more likely to be told by providers that their symptoms are emotionally driven rather than arising from a physical impairment.¹ Differences in doctors' evaluations of medical issues for male and female patients may be a key factor contributing to observed differences in treatment. Beyond impacting the treatments patients receive, medical evaluations also impact benefit eligibility in social insurance programs. Recent evidence suggests there are large gender disparities in social insurance programs that rely on medical evaluations. For example, Low and Pistaferri (2019) show that female applicants for Social Security Disability Insurance are 20 percentage points more likely to be rejected than similar male applicants. While there is mounting evidence that gender disparities exist in health care and related outcomes, there is little evidence on what drives these disparities and what policies may affect these disparities.

One frequently discussed policy prescription to address gender disparities in health care is diversifying the physician workforce. Despite gender parity in the training of new physicians, nearly two-thirds of current physicians are men (Kaiser Family Foundation 2019). In specialties that often evaluate physical impairments, like orthopedics, the gender imbalance is even larger (Jagsi et al. 2014). Some have pointed to gender imbalances among physician evaluators as a potential contributor to gender disparities in disability evaluations in social insurance programs. For example, a recent class-action lawsuit filed against the California workers' compensation system alleges that doctors are more likely to classify female occupational injuries as being non-work-related than male occupational injuries and argues that medical examiners being overwhelmingly male likely contributes to male-female disparities in benefit determinations.²

The gender of doctors could play a role in medical evaluations if patients and providers having similar characteristics and backgrounds improves provider-patient communication or reduces provider bias against a patient or against the types of health issues that the patient is likely to have. However, studying the role of provider gender in medical evaluations is difficult because both patients and doctors can often influence patient-doctor matches. In this paper, we overcome this key challenge by leveraging random assignment of doctors to patients for medical evaluations within the setting of workers' compensation insurance. To the best of our knowledge, our study provides the first estimates of the impact of the gender of the evaluating doctor on gender disparities in medical evaluations and subsequent social insurance benefit receipt. Further, we leverage additional administrative and survey data to provide evidence on potential mechanisms and to quantify gender homophily in patient-doctor matches in circumstances where patients have the ability to choose their own providers. Finally, we use our estimates to explore the policy implications of our findings.

Our study leverages comprehensive administrative data and random assignment of doctors to patients within the Texas workers' compensation insurance system. Random assignment of doctors to patients occurs in this setting through the dispute resolution process. Insurers and injured workers may request independent medical evaluations to settle disputes over an injured worker's impairment level, and these disputes are common, with more than a third of workers' compensation claims with cash benefits involving

¹For example, see Chen et al. (2008); Hoffmann and Tarzian (2001); Schulman et al. (1999); Tunks, Bellissimo, and Roy (1990).

²"Being a Woman Isn't a Pre-Existing Condition, Says This California Workers' Comp Suit", by Christina Cauterucci July 6, 2016 Slate.

a dispute-triggered independent medical evaluation. Medical doctors (MD), doctors of osteopathy (DO), and doctors of chiropractic (DC) can all apply through the Texas Department of Insurance to be included in the list of independent doctors eligible to perform these evaluations. Conditional on the injured worker's county, assignment of an injured worker to a doctor is random among doctors with the same credentials (i.e., MD, DO, and DC). Thus, after controlling for the injured worker's county and the doctor's credential, pre-determined characteristics of the injured worker should be unrelated to the gender of the evaluating doctor. The random assignment of doctors to patients means that differences in assessments between male and female doctors stem from the doctors themselves rather than from differences in the types of patients assigned to doctors.

An injured worker is eligible for continued cash benefits if the worker has an injury-related disability at the time of the independent exam, so a claimant will receive additional cash benefits after an exam if and only if the examining doctor assesses the worker as having a continued impairment. Our empirical approach compares the impact of being randomly assigned a female doctor rather than a male doctor on the receipt of subsequent benefits for female claimants to the analogous impact of the gender of the assigned doctor on the outcomes of male claimants. Our study draws on unique, linked administrative data covering all medical bills and cash benefits paid for workers' compensation claimants in Texas, as well as information on all medical evaluations performed by independent doctors through the dispute resolution system. To validate our reliance on the random assignment mechanism in this setting, we verify that doctor gender is uncorrelated with pre-determined patient and injury observable characteristics. We then estimate the differential impact of the gender of the assigned doctor on post-evaluation benefit receipt for exams for female patients versus exams for male patients.

There are several strengths of our research design and setting. First, our setting has random assignment of providers to patients, which is key for identifying the causal effect of being evaluated by a female doctor rather than a male doctor. Second, in addition to observing medical evaluations for female patients, we also observe evaluations for male patients, which means we can identify the effect of patient-provider gender match separately from any overall differences in evaluations performed by male and female doctors. Third, the number of randomized medical evaluations in our sample is large at 71,366 randomized medical evaluations performed by 1,461 doctors, and we observe rich demographic and medical data for all claimants undergoing these exams. Having rich baseline data from prior to the randomization allows us to verify that the gender of the assigned doctor is unrelated to baseline patient characteristics (as expected) and to explore heterogeneity. Fourth, the scope of these exams is set by state guidelines, and there is data on outcomes of these exams. While doctors' evaluations of patients have the potential to influence virtually all of patients' interactions with the health care system, studying factors that influence these evaluations is challenging because doctors' assessments of patients are typically difficult to measure, quantify, and isolate. Key features of this setting allow us to overcome these challenges—the standardized mandate of exams in this setting, the need to formalize assessments, and the data on later outcomes based on these exams. Fifth, these exams determine an important outcome for the patient: eligibility for subsequent cash disability benefits. Finally, in addition to having detailed administrative data on claims with randomized evaluations, we also have the same information for all other workers' compensation claims occurring during the study period, including those that never involve a dispute-triggered independent medical evaluation. Having data on all claims allows us to document gender gaps in benefit receipt more generally and to investigate how patients sort across providers when provider choice is available.

Beyond providing a useful context to examine the impact of patient-provider gender match on medi-

cal evaluations, independent medical evaluations are an important component of workers' compensation programs and are a topic of ongoing policy debate. Claim disputes are common in workers' compensation insurance, and many workers' compensation programs require independent medical evaluations to settle these disputes. Among injuries occurring in Texas in 2013, 39% of claims with cash benefits had a dispute that necessitated an independent medical exam, and these disputed claims accounted for a disproportionate share of claim costs—56% of program-wide claim costs or more than \$800 million per injury year. Similar rates of dispute-triggered independent medical exams are observed in other large workers' compensation programs, such as those in California and New York.³ Policies surrounding these exams—including rules for determining the pool of eligible doctors and the doctor assignment processes—are a subject of constant policy debate in workers' compensation programs (National Conference of State Legislatures 2010). Our study provides evidence on how policies related to these exams can affect claimant outcomes and gender disparities in these outcomes. In addition, the subject of the exams—evaluation of pain and disability—and gender disparities in these types of evaluations are of great policy interest in other related social insurance programs—including the Veterans Affairs disability program and Social Security Disability Insurance.⁴ More broadly, many social insurance and means-tested programs rely on program gatekeepers to determine benefit eligibility, and our findings point to the potential importance of gender match between program gatekeepers and claimants in other government programs.

Our estimates indicate that doctor-patient gender match increases evaluated disability and subsequent cash disability benefit receipt for female patients but has little impact on the outcomes of male patients.⁵ Compared to differences among their male patient counterparts, female patients randomly assigned a female doctor rather than a male doctor are 3.0 percentage points more likely to be evaluated as having an ongoing disability and receive 8.5% more cash benefits on average—or \$478 evaluated at the mean of \$5,627. There is no analogous gender-match effect for male patients. The estimates are very similar in alternative specifications that include doctor fixed effects rather than doctor gender fixed effects. Further, we note the magnitude of these effects is sizable. For instance, the 3.0 percentage point increase in the likelihood of being evaluated as disabled is nearly large enough to offset the entire observed gender gap in this outcome when male doctors evaluate claimants.

There may be several mechanisms that contribute to the importance of gender match in this setting. Differences in evaluations across male and female doctors may reflect provider discrimination against patients of the opposite gender or favoritism towards patients of the same gender. For instance, doctors may differentially empathize with patients of their own gender or discount medical conditions of patients of the opposite gender. Alternatively, differences in evaluations could arise if patients behave differently when matched with providers of the same gender, perhaps communicating more or different information that may affect their disability evaluation. Given our estimates indicate that gender match only matters for the evaluations of female patients, this suggests that mechanisms behind the importance of gender match do not operate symmetrically for men and women. While our setting does not allow us to precisely distinguish these potential mechanisms, we explore potential mechanisms through additional analysis within

³While we are aware of no national statistics on the use of independent medical evaluations in workers' compensation insurance programs, we examined their role in various specific state programs and found that they are ubiquitous. For example, in 2018, California and New York, which have more workers covered by workers' compensation insurance than all other states excluding Texas, had rates of more than 1 independent medical evaluation for every 3 claims.

⁴For examples, refer to Department of Veterans Affairs (2010); Murdoch et al. (2021); Trabanino (2020); Daly et al. (2013).

⁵We note that the effects we estimate may run through both gender and sex. Because gender and sex are highly correlated in practice and it is often unclear which is reflected in the data, we cannot distinguish the extent to which the effects we estimate reflect the importance of gender identity as opposed to biological sex. For simplicity and to follow conventions in the economics literature, throughout we use gender to refer to both gender and sex.

the setting of these randomized evaluations and through a supplemental survey we conduct to shed light on gender differences in patient-doctor interactions.

We explore whether the gender-match effect we document appears to arise from mechanisms inherently linked to gender rather than other patient or doctor characteristics that are correlated with gender. We begin by investigating whether female doctors (relative to their male counterparts) treat female patients differently because of some other patient characteristic correlated with gender. We find no significant differences in outcomes following exams by female and male doctors by other baseline injury or patient demographic characteristics—aside from patient gender. We also investigate whether female doctors treat female patients differently because of some other doctor characteristics correlated with doctor gender. To investigate this possibility, we regress post-exam patient benefit receipt on doctor \times female fixed effects to obtain doctor-specific measures of the female-male patient outcome differential. Correlating these fixed effects with doctor characteristics, we find that doctor gender stands out as the only doctor characteristic significantly associated with differential outcomes for female versus male patients. Collectively, this evidence suggests there is something fundamental about gender match between patients and providers impacting outcomes for female patients.

Heterogeneity analysis allows us to explore whether the gender-match effect is concentrated among particular subgroups, which could help shed light on potential mechanisms. The large number of randomized exams and rich baseline data allow us to explore heterogeneity in the effect by baseline patient demographic characteristics (income, age, marital status), injury characteristics (injury type, industry at time of injury, predicted severity), and characteristics of the patient's prior medical experience before the exam (prior doctor's gender, claimant contested case, medical spending on first day of injury). The gender-match effect for female patients appears across-the-board for exams with differing baseline patient demographic and injury characteristics, suggesting the gender match has near universal impacts on the outcomes of female patients.

Without an objective measure of the appropriate level of evaluated disability or the efficient level of cash benefits, we cannot determine whether female doctors are too generous or male doctors are too strict in their evaluations of female patients. While we do not have a comprehensive objective disability measure to benchmark evaluations against, we contextualize these findings using rich baseline data on patients to evaluate whether patient-doctor gender match ameliorates or magnifies differences in the evaluation of male and female patients conditional on observables. Controlling for all available baseline observable information about the patient, we see female doctors evaluate female and male patients as similarly disabled while male doctors evaluate female patients as less disabled than male patients. Specifically, the estimates indicate that having female doctor evaluators eliminates gender gaps in evaluated disability conditional on observables, in contrast to sizable gaps observed when male doctors evaluate patients. While this evidence is only suggestive, these results are consistent with male doctors evaluating female patients against a stricter standard than male patients and female doctors applying similar standards to male and female patients.

To complement our analysis of independent medical exams, we present novel evidence from a survey we conducted which suggests stark differences between men and women in their interactions within the health care system and suggests possible mechanisms behind the importance of patient-provider gender match in medical evaluations. The survey responses suggest that women—relative to men—more often report having a negative experience where a doctor didn't understand their concerns, had assumed something without asking, talked down to them, made them feel uncomfortable, or didn't believe them. When

asked about how a doctor's gender influences the likelihood of having a positive interaction, women were much more likely than men to report an own-gender doctor would be more likely to treat them with respect, understand their concerns, believe them, provide needed testing and treatments, make them feel comfortable, and ask appropriate questions instead of making assumptions. The asymmetry of responses between men and women supports the plausibility of the main finding from the randomized evaluations—that doctor gender is an important factor in the evaluation of female patients more so than in the evaluation of male patients—and suggests that differences in doctor behavior may be an important mechanism behind the main finding.

While our primary analysis focuses on disputed claims with randomly assigned evaluating doctors, we use comprehensive data on all workers' compensation claims in Texas to contextualize our primary findings. As observed among disputed claims, female workers' compensation claimants in the broader population are less likely to receive cash disability benefits than male claimants with the same observables. Controlling for a rich set of baseline claimant and injury characteristics, female workers' compensation claimants are 3.9 percentage points less likely to receive cash benefits than their male counterparts, which represents a 16.5% reduction relative to the mean rate of cash benefit receipt among male claimants.

Given evidence from our survey suggesting that women anticipate worse experiences when seeing male doctors rather than female doctors, women may express a preference for female doctors and may select female doctors more frequently when given the choice. We find evidence consistent with this, both in our analysis of the administrative data and in the survey we conducted. Specifically, we use the broader administrative data to investigate gender homophily in patient-doctor matches when patients can select their own doctor, which claimants in the Texas workers' compensation insurance system can do outside of the dispute resolution process. The estimates indicate female patients select a female treating doctor 1.6 percentage points—or 5.8%—more often than comparable male patients select a female treating doctor. While these estimates could reflect the preferences of both men and women to see own-gender providers, evidence from the survey we conducted suggests that women have stronger preferences to see own-gender providers. Responses to our survey indicate that women—compared to men—more often express a preference to see own-gender providers, more often select own-gender providers when given the choice, and report being willing to pay more to see an own-gender provider.

Finally, we explore the policy implications of our findings. One implication of our findings is that increasing gender diversity among doctors would increase evaluated disability and benefits for women, while having little impact on benefits for men. Our estimates imply that increasing the share of independent medical evaluations performed by female doctors from 17% to 50% would cause a 0.86 percentage point increase in the share of female patients evaluated as disabled, closing 39.8% of the gender gap conditional on observables among disputed claims. An alternative random assignment mechanism that randomly assigns doctors of the same-gender to evaluate disputed claims—rather than the current gender-blind random assignment process—would increase the share of female patients who receive cash benefits by 2.6 percentage points, offsetting the entire gender gap conditional on observables among disputed claims.

We also consider the potential policy implications of our findings for gender gaps in workers' compensation insurance more broadly, beyond disputed claims. Specifically, we use our estimates to characterize the potential impact of two types of policies: increasing the share of female doctors and increasing relative gender homophily in patient-doctor matches. Holding all else fixed, this analysis suggests that increasing the share of female treating doctors by 21.6 percentage points—moving from 28.4% to parity—would lead to a 31.6% decrease in the observed gender gap. Alternatively, if we hold fixed the overall gender com-

position of treating doctors, this analysis suggests that an increase in relative gender homophily from the observed level to two (i.e., where female patients select female doctors at twice the rate that male patients select female doctors), would lead to a 16.6% decrease in the gender gap conditional on observables. More broadly, we use our estimates to characterize the mix of gender diversity among doctors and relative gender homophily in patient-provider matches that would lead to a given decrease in the gender gap conditional on observables. Overall, this analysis reveals that these types of policies may have substantial impacts and complement one another in closing gender gaps in evaluated disability.

The remainder of the paper proceeds as follows. Section 2 describes related work and our contribution. Section 3 describes the setting and the data, and Section 4 outlines the empirical strategy and identifying variation. Section 5 presents the main estimates of the effects of evaluations by doctor gender, leveraging disputed claims with randomized evaluations. Section 6 provides broader evidence on gender gaps in evaluated disability and gender homophily in patient-provider matches. Section 7 describes results of the survey we conducted to shed light on potential mechanisms. Section 8 conducts back-of-the-envelope policy counterfactual analysis, and Section 9 concludes.

2 Related Literature

This paper contributes to several literatures. This paper is motivated by a recent literature documenting gender gaps in health care treatments and related outcomes. A number of recent medical studies have documented that female patients are less likely to receive aggressive treatment for a range of medical conditions including heart-related issues (Hernandez et al. 2007; Pelletier et al. 2014; Vaccarino et al. 2005) and blood clots (Pietropaoli et al. 2010).^{6,7} Among patients reporting similar pain levels, females are less likely to receive treatment to alleviate pain (Chen et al. 2008). The observed gender disparities in medical treatments may arise if doctors interpret symptoms differently for male and female patients. Some recent evidence suggests this may be the case, with studies documenting that females reporting pain are more likely to be told that their pain is emotionally driven rather than arising from physical sources (Côté and Coutu 2010; Hoffmann and Tarzian 2001; Samulowitz et al. 2018).⁸ Besides impacting the treatments patients receive, differences in the way medical professionals evaluate men and women may result in gender disparities in social insurance benefits that depend on medical evaluations. Recent work suggests gender disparities in social insurance benefit receipt are large, with Low and Pistaferri (2019) documenting that women are 20 percentage points more likely to be denied Social Security Disability Insurance benefits than men with the same observable characteristics. Despite growing evidence of gender disparities in health care and related settings, it is unclear what drives these gender differences and what interventions may affect these disparities.⁹ The present paper moves this literature forward in several ways. We use administrative claims data to

⁶In other related work, Chandra and Staiger (2010) document racial and gender gaps in heart attack treatments, and they find evidence consistent with lower benefits of treatment for women and blacks, pointing to the potential importance of statistical discrimination in driving gaps in heart attack treatments.

⁷Dupas and Jain (2021) illustrate there are large gender disparities in health care treatments in India due to gender differences in the propensity to seek health care, where the uneven allocation of resources within families may be a primary channel for observed gender differences in patients seeking care in the developing country context.

⁸See Cabral and Dillender (2020) for a more comprehensive review of this literature.

⁹While observational studies point to large gender differences in the evaluation of pain and disability, evidence from vignettes is more mixed (Low and Pistaferri 2019; Weisse, Sorum, and Dominguez 2003). We note an important limitation of vignette studies is that they are unable to capture features of real-world interactions that may influence exam outcomes. For instance, in-person interactions during medical exams may influence provider assessments about the credibility of the patient's self-reported symptoms or provider empathy toward the patient. Given the limits of both observational and vignette studies, it is unclear whether gender disparities in the evaluation of pain and disability are due to differences in unobservable patient characteristics or some systematic feature of the health care system or disability evaluation system. Our results highlight that the gender of the evaluating doctor is an important determinant of evaluation outcomes for women, and thus a key feature of the health care system—the gender imbalance in the provider workforce

document substantial gender gaps in medical evaluations, conditional on a rich set of baseline observable characteristics. Further, we leverage random assignment of doctors to patients to provide the first evidence on the role of doctors—and in particular, the role of doctor-patient gender match—in contributing to gender disparities in medical evaluations and linked social insurance benefits. The evidence indicates that gender match between patients and doctors increases evaluated disability and cash disability benefits for women and works to nearly close gender gaps conditional on observables in this setting. Our findings suggest that the under-representation of female doctors may play a large role in driving observed gender disparities in medical evaluations and related outcomes.

Our work also contributes to an emerging literature considering the impacts of race or gender match between providers and patients in health care settings. A few recent studies document correlations that suggest that gender or race concordance between patients and providers may reduce hospital mortality (Greenwood, Carnahan, and Huang 2018; Hill, Jones, and Woodworth 2020). Similarly, recent experimental work shows that black men take up preventive care at higher rates when counseled by black doctors rather than by white doctors (Alsan, Garrick, and Graziani 2019). Our paper extends this literature in several ways. We leverage random assignment to provide the first estimates of the causal effect of gender match between patients and doctors on medical evaluations and subsequent social insurance benefit receipt. In doing so, our study provides the first evidence on the importance of demographic concordance between patients and doctors in medical evaluations—the key first step in most health care interactions.¹⁰ Further, several features of our setting—for instance, the random assignment of doctors, a diverse pool of patients and doctors, the large scale of the randomized evaluations, the availability of broader administrative data—allow for a more transparent and comprehensive evaluation of the effects of gender match on medical evaluations compared to prior work investigating the effects of gender match or race match on other health care outcomes. For instance, the random assignment of doctors allows us to overcome identification challenges usually present when patients select their own providers, and the availability of data on both male and female patients allows us to control for doctor fixed effects and to estimate impacts on outcomes of both male and female patients (and hence impacts on the gender gap in this setting). Further, administrative data from the broader setting beyond the randomized evaluations allow us to provide context to our estimates and conduct policy counterfactuals evaluating the effects of diversifying the doctor workforce. Our evidence on gender homophily in patient selections of providers in broader settings complements prior work illustrating gender homophily in patient selections of urologists (McDevitt and Roberts 2014) and in referrals among doctors (Zeltzer 2020). Further, our evidence on gender disparities in evaluations of patients complements work by Sarsons (2019), which documents gender disparities in how physician ability is interpreted following positive and negative patient outcomes.

This paper also provides evidence on the importance of gatekeeper discretion in social insurance and social safety net programs. Many of the largest public programs have a gatekeeper model, where designated gatekeepers are charged with evaluating benefit eligibility among program claimants. For instance, medical impairments and the ability to work are assessed by doctors in workers' compensation insurance

—may be an important explanation for observed gender disparities in medical evaluations and related outcomes.

¹⁰Health care decisions typically involve complex interactions between patients and doctors, where the degree to which health care decisions are determined by patients versus doctors depends on the type of health care interaction. Our work provides the first evidence on the role of demographic concordance between doctors and patients in doctors' evaluations of patients' medical conditions—the first step in most health care interactions and an aspect of health care interactions where doctor agency is central. In this way, our work complements prior work by Alsan, Garrick, and Graziani (2019), which investigates the role of doctors' race in black patients' decisions to take up effective and broadly recommended preventive care—a setting where patient agency is central. See Section 5 for discussion of potential mechanisms behind the importance of demographic concordance in our setting and health care interactions more broadly.

and by disability examiners in Social Security Disability Insurance; caseworkers interview applicants to assess eligibility for the federal Supplemental Nutritional Assistance Program; judges determine access to Chapter 13 bankruptcy protection; hearing officers settle disputes over unemployment insurance eligibility; and interviewers assess eligibility for the Temporary Assistance for Needy Families program. Prior work in the economics literature has recognized that gatekeepers may exercise discretion in eligibility determinations, with papers leveraging variation in gatekeeper leniency as an instrument to look at the effect of benefit receipt on later outcomes (e.g., Maestas, Mullen, and Strand 2013; Dobbie and Song 2015). However, little is known about what factors drive gatekeeper discretion or about the distributional impacts of gatekeeper discretion on program benefits. This is an important gap in the literature given that hundreds of billions of dollars in public program benefits are awarded at the discretion of gatekeepers each year and that gatekeeper discretion has the potential to have important distributional impacts on the receipt of program benefits and on broader income disparities. The findings from our paper suggest gender match between gatekeepers and program claimants can meaningfully affect the benefits women receive in social insurance programs and can work to close gender disparities in benefits. Further, our findings highlight how the design of processes by which gatekeepers are assigned to claimants can impact outcomes and disparities.

Finally, our findings complement a larger literature investigating the impact of the gender of authority figures on outcomes for women in contexts ranging from education to the labor market. Studies have documented that females are more likely to pursue STEM-focused education when matched with female teachers (Lim and Meer 2019) or female professors (Carrell, Page, and West 2010). Evidence of the short-run effects gender match has on student performance is more mixed, with some studies finding improvements when female students are matched with female teachers (e.g., Muralidharan and Sheth 2016; Lim and Meer 2017), while others have failed to find effects (e.g., Antecol, Eren, and Ozbeklik 2015). A related literature explores these issues in the context of the impact of firm leadership on women's career outcomes. The results in this literature are mixed, with some studies finding no impact of gender quotas for firm governing boards on outcomes of women at the associated firms (Bertrand et al. 2018; Maida and Weber 2019), while other studies find that more direct female managers have a positive impact on outcomes for women (e.g., Biasi and Sarsons 2020; Cullen and Perez-Truglia 2019; Casarico and Lattanzio 2019). Related work on the academic publication process has found no impact of gender match on evaluations of peer-reviewers (Abrevaya and Hamermesh 2012; Card et al. 2019). Our work complements this literature by providing evidence on the role of evaluator gender in the context of doctors performing medical evaluations for work-limiting disabilities. We find that having a female doctor evaluator substantially impacts outcomes for women, while the gender of the evaluating doctor has little impact on outcomes for men. While gender-match effects in educational and professional settings may run through a complex set of mechanisms (e.g., role model effects, differing management or teaching styles, networking, mentoring, discrimination, etc.), the authority figures in our setting (doctors) have a relatively narrow mandate—to medically evaluate a patient's ongoing degree of disability—limiting the set of the plausible mechanisms underlying effects in this setting. Our findings may point to the broader importance of gender match in settings where authority figures are charged with evaluating women.

3 Background and Data

3.1 Workers' Compensation Insurance

Workers' compensation is a state-regulated insurance system that provides covered employees with cash and medical benefits for work-related injuries or illnesses. Workers' compensation insurance provides cov-

erage regardless of whether the employer or employee is at fault for the workplace injury.¹¹ While there is variation across states in the details of the workers' compensation insurance system, the basic structure of workers' compensation is similar across states. Workers' compensation insurance is the first payer for injury-related medical expenditures and covers all injury-related medical spending at no out-of-pocket cost to the claimant, regardless of a claimant's work status or receipt of cash benefits. Temporary income benefits replace wages that claimants forgo because of missed work that arises from an on-the-job injury and are the most common type of cash benefits in workers' compensation. Permanent impairment benefits are given to workers with a permanent impairment from an on-the-job injury and are the second most common type of cash benefits. Workers' compensation insurance provides additional benefits in some specified circumstances, such as supplemental benefits for claimants with a rare set of very severe permanent impairments (e.g., lost use of both eyes) or death and burial benefits to the surviving family of those who die on the job.

A key feature of workers' compensation insurance is that medical providers evaluate impairments and determine eligibility for cash benefits. In Texas, as in many states, the delivery of medical care and the oversight of cash benefit eligibility in workers' compensation insurance follow a "gatekeeper" model. Workers' compensation claimants choose a "treating doctor", who is responsible for overseeing the claimant's medical care, evaluating the claimant's medical improvement, and assessing any permanent impairment the claimant may have.¹² In addition to receiving reimbursement for typical procedures billed by providers, doctors treating workers' compensation claimants receive payments for additional "case management services" that pertain to their particular role in overseeing the medical care and cash benefit eligibility of injured workers.

There are two main types of cash benefits: temporary income benefits and permanent impairment benefits. In Texas, injured workers are eligible for temporary income benefits after a waiting period of seven days. Temporary income benefits are terminated at the earliest of the following: (i) the employee decides to return to work, (ii) the employee's doctor determines that the employee has reached "maximum medical improvement", and (iii) the maximum duration is met. The maximum temporary income benefits duration in Texas is two years (104 weeks), but workers rarely receive temporary income benefits for two years. Temporary income benefits replace 70% of a claimant's prior average weekly wage, subject to a maximum and minimum weekly benefit level.

Once workers have reached maximum medical improvement or have exhausted their temporary income benefits, they are eligible for permanent impairment benefits if they are still impaired. Treating doctors who conclude that a claimant has a permanent impairment assign an impairment rating—the percentage of permanent impairment of the whole body resulting from the current compensable injury. The worker is then paid an unconditional cash transfer that is a function of the severity rating of his permanent impairment and his prior average weekly wage. In Texas—as in many states—permanent impairment benefits are calculated as the product of the claimant's permanent impairment severity rating and the eligible wage replacement rate, where the replacement rate is linear in the claimant's prior average weekly earnings subject to a maximum which is binding for at a considerably lower level of prior earnings than the temporary income benefits maximum.

¹¹Workers' compensation insurance serves as the exclusive legal remedy for covered workers for workplace injuries, meaning that injured workers cannot sue their employers for negligence.

¹²About half of workers' compensation claimants in Texas receive treatment through provider networks maintained by insurers (TDI 2020). For these workers, the choice of treating doctor must come from the insurer's network.

3.2 Independent Medical Evaluations

While the treating doctor typically serves as the gatekeeper for cash benefits, insurers and/or injured workers may disagree with the treating doctor's assessment of a claimant's impairment level. As with most states, Texas maintains a database of independent doctors—referred to as “designated doctors”—who can resolve these disputes by performing independent medical evaluations. Either an insurer or an injured worker can request an independent evaluation if they disagree with the treating doctor's assessment, though insurers are much more likely to request them. For injuries occurring from 2013 to 2017, 87% of designated doctor exams in Texas were requested by insurers. These designated doctors are most commonly tasked with settling disputes about whether a claimant has reached his/her maximum medical improvement and the degree of the permanent impairment resulting from the injury (if any).

We focus on independent exams where the designated doctors have been asked to assess the workers' maximum medical improvement and/or permanent impairments. Nearly all injuries (99%) that go through the designated doctor process during our period of analysis meet this criteria.¹³ In principle, a designated doctor could be asked to assess maximum medical improvement and permanent impairment severity separately. However, because permanent impairments are assessed once maximum medical improvement has been reached, designated doctors are nearly always (> 99.8% of the time) asked to assess both dimensions in a single examination if either determination is requested, though the request for the permanent impairment determination is only relevant if the designated doctor assesses the injured worker as having reached maximum medical improvement.

Medical doctors (MD), doctors of osteopathy (DO), and doctors of chiropractic (DC) can apply to be designated doctors through the Texas Department of Insurance. Their reports are given presumptive weight in dispute resolution, and insurance carriers are required to pay income and medical benefits based on the opinion of the designated doctor. Designated doctors cannot have a conflict of interest with the insurer or patient associated with an assigned claim. For example, designated doctors cannot have treated the patient prior to the independent exam and cannot have a contract with the insurer or network associated with the claim. Once a designated doctor has been assigned to a claim, that doctor is the designated doctor for the initially requested independent medical exam and for any subsequently requested independent medical exams for that claim.

The role of the designated doctor exam is narrow and well-defined. The designated doctor is called upon by the Texas Department of Insurance to serve as an “impartial, objective medical expert” to settle specified claim-related disputes. Designated doctors do not have previous or ongoing relationships with patients for these disputed claims outside of independent medical exams, and they are not allowed to recommend or provide treatment to these patients. To become certified as a designated doctor, doctors need to complete required training and testing and must re-certify every two years with the Texas Department of Insurance. Because the typical designated doctor performs a relatively small number of these exams annually (median 12 and mean 23.7 for doctors performing at least one exam), designated doctor exams make up a small share of any given doctor's business and these doctors likely spend most of the professional time and effort treating patients in their standard practice.

Texas law defines the scope of designated doctor evaluations, the information doctors are allowed to consider, and the requirements for designated doctors when reporting and justifying their decisions.

¹³Though it is rare, designated doctors can be asked to settle other types of disputes besides disputes about maximum medical improvement and permanent impairments, such as whether the impairment was caused by a compensable injury and the extent of the worker's compensable injury.

Designated doctors must review a claimant’s medical records—which are required to be supplied by the claimant’s treating doctor and insurer—prior to the designated doctor exam.¹⁴ The designated doctor must also review any medical records and medical history information supplied by the claimant. At the designated doctor exam, the doctor conducts a physical examination of the claimant—often including range of motion testing and tests specific to the affected body system(s). The designated doctor then compares the information gathered through the physical examination and medical records with specified guidelines for these evaluations.¹⁵ In reporting the results of the exam, the designated doctor is required to justify his or her decision based on the evidence gathered and specified medical guidelines.¹⁶

The assignment of claimants to designated doctors is designed to be random within a local area, among doctors eligible to perform an exam. Once an independent evaluation has been requested, the claimant is assigned to the next available designated doctor within the claimant’s county who: (i) has the appropriate credential to evaluate the injury and (ii) does not have a conflict of interest. Any designated doctor (MD, DO, or DC) can evaluate musculoskeletal injuries, while only MDs and DOs are authorized to evaluate certain non-musculoskeletal injuries, like mental and behavioral disorders.¹⁷ As the vast majority of workers’ compensation claims and independent exams are for musculoskeletal injuries, all providers are eligible to do most exams.

Since insurers and claimants have no way of knowing who is the next available designated doctor, they cannot time their requests so that they have a specific designated doctor. Thus, after controlling for the claimant’s county and the designated doctor’s credential, the designated doctor assigned to a claimant is random from among designated doctors with that credential in the claimant’s county. Prior to 2013, a representative at the Texas Department of Insurance assigned claimants to providers. But to prevent the possibility of manipulation and to ensure that process approximates random assignment as designed, the Texas Department of Insurance implemented an automated system to assign claimants to designated doctors as of 2013. The Texas Department of Insurance maintains a confidential list of certified designated doctors in each county. The automated system cycles through this list, assigning the next available designated doctor in the claimant’s county who is qualified to perform the assessment (i.e., has no conflicts of interest and has appropriate credentials) and designated doctors move to the bottom of the list upon assignment. This automated assignment process is confidential, and insurers and claimants do not observe the order of—or current position within—the list of certified designated doctors in each county.

As we describe further in Section 4, our estimation approach controls for the credential of the assigned designated doctor (known ex post) by claimant county fixed effects, as the assignment process means the designated doctor assigned to a claimant is random among designated doctors with that credential in the claimant’s county. An alternative approach to isolate conditional random assignment would be to instead control for each injury type by county combination—the ex ante information the regulator uses to decide

¹⁴Texas law requires a claimant’s prior medical records to be supplied to the designated doctor by the claimant’s treating doctor and insurer at least three days before the designated doctor exam (28 TAC §127.10(a)). See 28 TAC §127.10 for more information on designated doctor exam requirements.

¹⁵Texas law states that designated doctors will apply when appropriate the American Medical Association Guides for the Evaluation of Permanent Impairment and Texas Department of Workers’ Compensation return-to-work guidelines. Designated doctors will also consider Texas Department of Workers’ Compensation treatment guidelines and other evidence-based medical guidelines when appropriate (28 TAC §127.200(a)(7)).

¹⁶See 28 TAC §127.200(a) for more detail.

¹⁷As discussed further below, our research design will isolate as good as random assignment of doctors to claimants conditional on claimant county and provider credential (MD, DO, or DC). For this reason, we exclude less than 3% of designated doctor exams for claimants with specified diagnoses that the Texas Department of Insurance flags as requiring an evaluation by an MD or DO with a specific sub-specialty. Certain rare complex injuries, such as traumatic brain injuries or spinal cord injuries, require an evaluation by an MD or DO with certain board-certified sub-specialties. Refer to 28 TAC §127.130(b) for more detail. We use diagnosis codes from the designated doctor exam to identify and exclude claimants whose exams potentially require specialized expertise.

which subset of designated doctors within a county are eligible to evaluate a claimant. A potential concern with basing the estimation on a control for injury type is that we can only imperfectly observe the injury type information used by the regulator in selecting which doctor credentials are required to evaluate a claimant.¹⁸ Thus, to be conservative, our baseline approach is to control for credential of the assigned doctor by county, as the assignment mechanism means that it is as good as random which doctor was assigned to the claimant among designated doctors in that county with that credential. In practice, we obtain very similar estimates if we control for injury type by county fixed effects instead of credential by county fixed effects in our analysis.¹⁹

Claimants will receive cash benefits after a designated doctor visit if and only if the designated doctor assesses the claimant as having an impairment on the day of the evaluation. Conditional on the designated doctor viewing the claimant as being impaired, the type of benefits that claimants will receive depends on the doctor's assessment of the injury's healing trajectory. If the designated doctor decides that a claimant is impaired but that the claimant has reached maximum medical improvement, the designated doctor will assign the claimant an impairment rating, and the claimant will receive additional permanent impairment cash benefits based on that rating. If the designated doctor decides that a claimant is impaired and still healing from an injury, the claimant can receive additional temporary income benefits and can then be assessed for a permanent impairment at a later date, either by the designated doctor or by the claimant's treating doctor.

For claimants assessed as having a continued impairment, designated doctors can decide the amount of future benefits at the initial visit by assigning a permanent impairment rating or by assigning a future date at which maximum medical improvement will be reached. If designated doctors assess a claimant as having a continued impairment that is still healing and decline to provide a future date at which maximum medical improvement will occur, decisions about the amount of future benefits revert to the claimant's treating doctor, unless and until the insurer or claimant requests another designated doctor evaluation. Since treating doctors' benefit decisions and insurers' (and workers') decisions about requesting further independent evaluations may be influenced by the decisions that the designated doctor makes in the initial evaluation, we focus on all benefits received after the initial evaluation.

3.3 Data and Sample

To consider the role of provider's gender in claimants' outcomes, we have compiled a unique administrative dataset through open records requests submitted to the Texas Department of Insurance (TDI). As part of an open records request, we received data on all independent medical evaluations that occurred from 2005 to 2017. Since a lack of influence on the designated doctor assignment process is crucial to the empirical strategy, we focus on designated doctor evaluations for injuries that occurred from 2013 to 2017 so that claimants are assigned to a designated doctor through the automated system (TDI 2018a). We focus on benefits received within 12 months of the exam date and choose 2017 as the exam year cutoff so that we have information on claimants' benefit receipt for at least one year after designated doctor exams in the

¹⁸Our measure of injury type, which is based on the claimant's diagnosis on the first day of treatment for that injury, is imperfect as it could in principle miss aspects of a claimant's injury that the regulator considers when deciding which types of doctors are eligible to evaluate the claimant (MDs, DOs, and/or DCs). For example, while it is straightforward to identify musculoskeletal injuries, our measure of injury type may miss secondary non-musculoskeletal conditions (e.g., subsequent secondary diagnoses of mental health conditions, secondarily affected body systems) that may lead the regulator to require an evaluation by an MD or DO, rather than a DC.

¹⁹See Appendix Table A2 for estimates from this alternative specification. It is not surprising that we obtain similar results using either approach. The measurement error in our injury type variable may be limited in practice, and the vast majority of exams are for claimants with musculoskeletal injuries, whose exams can be performed by any designated doctor regardless of credential.

sample. Since subsequent designated doctor exams are endogenous to the first designated doctor exam, we focus on the date of the first independent medical evaluation. In addition to conducting analysis that focuses on the sample of claimants with randomized medical evaluations, we also use the full sample of claims from 2013 to 2017 to study gender differences in benefit receipt and in the choice of treating doctors more broadly. We exclude the 3.6% of claimants with missing gender information from all analyses.

We combine the information on designated doctor visits with data on the cash and medical benefits that claimants receive (TDI 2018*b*). The cash benefit data include claimant-level information on type of cash benefits received, prior average weekly wage, total benefits received, benefit replacement rate, benefit start and end dates, and injury date (month-year). The medical benefit data come from all medical, pharmacy, and case management bills paid for by workers' compensation insurance and include the following information: procedure type (CPT codes), amount paid, amount charged, diagnoses (ICD-9 codes), date, place of service, and provider information. The data also include gender, birth date (month-year), and zip code.

For claimants who receive cash benefits, the data also include information on pre-injury weekly wages and industry. We present summary statistics for this information later in this section, and we also use this information for some supplemental analysis on potential mechanisms in Section 5. While wage and industry information is useful for describing the sample and providing context for our estimates, it is important to note that exam outcomes can influence whether this information is present in the data. Thus, our primary analysis does not use information on pre-injury wages or industry, and supplemental analysis using this information should be interpreted with caution.

To determine the gender of the designated doctor, we merge data on providers from the Centers for Medicare & Medicaid Services' (CMS) National Provider Identifier (NPI) registry based on the NPI variable in the administrative data (CMS 2019). We focus on evaluations for which we know the gender of both the provider and the claimant. Our final sample has 71,366 designated doctor exams performed by 1,461 designated doctors. We compute doctors' years of experience using information on graduation and birth years from the Medicare Physician Compare File (CMS 2021) and from an open records request to the Texas Board of Chiropractic Examiners (Texas Board of Chiropractic Examiners 2020). The Medicare Physician Compare File also provides information on the medical schools of MDs and DOs.

We focus on two measures of benefit receipt after a designated doctor exam. The first is an indicator variable equal to one if the claimant receives any cash benefits in the year after the exam. As described above, this variable equals one if and only if the doctor assesses an injured worker as being disabled at the time of the exam. To assess the value of benefits received after the exam, we create a second variable—"normalized additional benefits"—to reflect the total cash benefits received in the year after the exam valued at the mean benefit rate for each benefit type based the population-wide distribution of pre-injury wages. To create this measure, we first calculate the mean benefit rate for each benefit type—the population-wide mean inflation-adjusted benefits paid per week receiving benefits based on all claims with nonzero benefits.²⁰ We then calculate the normalized additional benefits received as the sum of the post-evaluation temporary income benefits and permanent impairment benefits valued at the mean benefit rate for that benefit type. Specifically, we multiply the mean benefit rate for each type of benefit by the number of weeks

²⁰As described above, temporary income benefits are paid for the weeks that individuals miss work while healing from temporary impairments. Permanent impairment severity—as rated by a claimant's treating doctor or designated doctor—determines the number of weeks permanent impairment benefits are paid, though these benefits are paid regardless of whether a claimant remains out of work or returns to work. Among individuals who receive non-zero benefits for a particular benefit type, we calculate an individual's inflation-adjusted benefit rate for that benefit type as the ratio of the inflation-adjusted dollars of benefits received to the weeks receiving benefits. We then calculate the population-wide mean weekly rate for a particular benefit type by taking the mean weekly rate among individuals who receive non-zero benefits for that benefit type.

the claimant received benefits and then add these two amounts to calculate the normalized additional cash benefits. A feature of this measure is that it depends on only the claimant's evaluated degree of disability rather than the claimant's pre-injury wages.

Table 1 displays means of key variables for male and female claimants, both for the full set of claimants with injuries occurring from 2013 to 2017 and for the subset of claimants who received independent medical evaluations by 2017.²¹ In the sample of claimants who received independent medical evaluations, male claimants have higher mean first-day and three-month medical spending than females do. Male claimants are also more likely to first receive care in the emergency department than female claimants are and receive income and impairment benefits at higher rates than female claimants do. Female claimants are older on average, and muscle issues and sprains account for higher shares of claims for females than for males.

Our estimates of the gender-match effect provide direct evidence related to outcomes among claims that receive independent medical evaluations. We note that this is a large group of claims and one that makes up a disproportionate share of workers' compensation costs—39% of all workers compensation claims with cash benefits involve a dispute-triggered independent medical evaluation and 56% of workers' compensation claim costs are attributable to such claims.²² Independent medical exams are central to resolving workers' compensation disputes in Texas and beyond, and processes surrounding these exams (e.g., the examiner assignment process) are a constant topic of policy debate. While these exams are of policy interest themselves, we may also be interested in the extent to which the gender-match effects we document may extend to broader populations—workers' compensation claimants more generally and other contexts with medical evaluations (e.g., other health care settings, other settings with disability evaluations). While we cannot directly assess the wider applicability of our findings, we use broader administrative data on all workers' compensation claims to provide some context on how the subset of claims with independent medical exams during our analysis period differ from workers' compensation claims more broadly and how gender differences compare across these samples.

There are some notable differences between claims that receive independent medical evaluations and the full sample of claims holding gender constant. Within gender, claimants receiving independent medical evaluations are older, are more likely to first receive care in the emergency department, and have higher medical spending and benefit receipt relative to the full sample of claimants. Claims for muscle issues and sprains make up a higher share of claims receiving independent medical evaluations than they do for the full set of claims. While there are notable differences in the types of claims that receive independent medical evaluations, we observe similar gender differences among the full sample of claimants as we do among the sample of claims with independent medical evaluations. Compared to female claimants, male claimants have higher initial medical spending for their injuries, are more likely to first receive care in the emergency department, receive cash benefits at higher rates, are younger, and have a lower share of claims due to

²¹Note that while 6.6% of claims for injuries occurring from 2013 to 2017 had an independent medical exam by the end of 2017, a higher share of these claims will have an exam at some point, since claims are more likely to have an independent medical exam the longer they are open. For example, for claims for 2013 injuries, 8.8% have had an independent medical exam by the end of the sample period. Because independent medical exams typically settle disputes about the worker's continued inability to work, they predominately occur for claims with cash disability benefits, which have higher stakes for both employers and injured workers than claims involving only medical benefits. Among claims for 2013 injuries, the 22% of claims receiving any cash benefits accounted for 86% of total medical and cash benefits paid out through the end of the sample period and 39% of these claims had an independent medical exam by the end of the sample period.

²²For these calculations, we focus on all workers' compensation claims for injuries occurring in 2013. We first calculate each injury's total medical and cash benefits through 2017. Among 2013 injuries, we identify those that had ever had an independent exam at any point between 2013 and 2017. If an independent medical exam were to take place for a 2013 injury, we would expect that exam to occur by 2017 given that temporary income benefits are capped at two years. Finally, we calculate the share of overall costs for 2013 injuries that comes from injuries with an independent medical exam.

muscle issues and sprains. Beyond the similar gender differences on observable across these samples, our analysis in Sections 5 and 6 document large gaps in cash disability benefit receipt in both the full sample and the independent medical exam sample. While the patterns of gender differences in observables and cash benefit receipt (conditional on observables) are broadly similar across these samples, it is important to emphasize that appropriate caution should be used in extrapolation of our evidence on gender-match effects from the independent medical evaluation sample to broader contexts. For instance, when we evaluate policy counterfactuals, we make the distinction between counterfactuals within the independent medical evaluation sample—pertaining to the pool of evaluating doctors or the evaluator assignment process—and broader counterfactuals related to evaluations of workers’ compensation claimants in general. We emphasize that the latter set of counterfactuals require extrapolation beyond the identifying variation, and thus should be interpreted with more caution.

Table 2 displays characteristics of doctors performing evaluations in Texas and more broadly. Column 1 of Table 2 displays characteristics of the 1,461 designated doctors who perform independent medical exams during our analysis period, while column 2 shows characteristics of the broader set of doctors treating patients within the Texas workers’ compensation insurance system. There are a few notable patterns comparing these two columns. The set of designated doctors performing independent medical evaluations makes up more than 26% of all doctors treating workers’ compensation patients and more than 2% of all doctors practicing in Texas more broadly (those described in column 3). Second, the share of doctors who are female is similar among designated doctors and all doctors treating workers’ compensation patients. Third, relative to the population of doctors treating workers’ compensation patients, a larger share of designated doctors have a DC credential rather than an MD or DO credential. Finally, among doctors with MD or DO credentials, the distribution of specialties and the share who graduated from a Top 25 medical school are broadly similar among designated doctors and doctors treating workers’ compensation patients.

Columns 3 and 4 display characteristics of all doctors in Texas and all doctors in the United States. Relative to doctors overall in Texas, a larger share of doctors treating workers’ compensation patients hold DC credentials and specialize in orthopedics or internal/family medicine among those with MD or DO credentials. Among those with an MD or DO credential, doctors treating workers’ compensation patients in Texas and doctors overall in Texas are similarly likely to have graduated from a Top 25 medical school. Compared to doctors treating workers’ compensation patients, the shares of doctors who are female are also higher in columns 3 and 4, though there are still about twice as many male doctors as female doctors. Along the dimensions considered, Texas doctors are similar to doctors nationally.

4 Empirical Strategy

4.1 Estimating Equations

Our empirical strategy takes advantage of random assignment of designated doctors to claimants. Conditional on the designated doctor’s credential and the claimant’s county, the assignment of designated doctors to claimants is random, and thus a claimant’s underlying characteristics should not be related to the gender of the designated doctor. A key advantage of our setting is that we observe evaluations for claimants of both genders. This allows us to estimate and control for any cross-the-board differences in the evaluations of male and female doctors.

Our baseline specification is as follows:

$$y_{icp} = \beta_1 female_claimant_i + \beta_2 female_doctor_p + \beta_3 female_claimant_i * female_doctor_p + \theta_{cp} + f(\mathbf{X}_{icp}) + \epsilon_{icp}, \quad (1)$$

where i indexes the claimant, c indexes the claimant’s county, and p indexes the provider. In this specification, y_{icp} represents the dependent variable, $female_claimant_i$ is an indicator for the claimant being female, $female_doctor_p$ is an indicator for the designated doctor performing the exam being a female, θ_{cp} is a vector of fixed effects for a claimant’s county by the doctor’s credential, and \mathbf{X}_{icp} is a set of additional controls included in some specifications that describe characteristics of the claimant, provider, and initial evaluation. In our baseline specification, \mathbf{X}_{icp} includes controls for the year of the claimant’s injury and the year of the exam. Standard errors are clustered at the doctor level.

Within this specification, β_2 represents the difference in outcomes among male claimants randomly assigned to a female doctor relative to those assigned to a male doctor, while $\beta_2 + \beta_3$ represents the analogous difference in outcomes among female claimants. Comparing these differentials across claimants, we can interpret β_3 —the main coefficient of interest— as representing the differential impact of the assigned doctor’s gender on outcomes of female patients relative to male patients. While our discussion largely focuses on the estimates of β_3 , we also discuss estimates of β_2 as these capture whether female doctors are systematically more or less generous in injury assessments across-the-board (for all claimants) compared to their male counterparts. We also report β_1 throughout, so we can contextualize the main estimates relative to gender gaps in benefit receipt among claimants evaluated by male designated doctors.

In addition to the baseline specification above, we also estimate two variants. First, we estimate a specification including doctor fixed effects:

$$y_{icp} = \delta_1 female_claimant_i + \delta_2 female_claimant_i * female_doctor_p + \psi_{cp} + \gamma_p + h(\mathbf{X}_{icp}) + e_{icp}, \quad (2)$$

where γ_p is a vector of doctor fixed effects. By including doctor fixed effects, the identification of the key parameter of interest in this specification (δ_2) comes from comparing outcomes within doctor for randomly assigned male and female claimants. Second, we also estimate a specification focusing on only female claimants for comparison:

$$y_{icp} = \alpha female_doctor_p + \omega_{cp} + g(\mathbf{X}_{icp}) + \mu_{icp}, \quad (3)$$

where the coefficient α measures the differences in outcomes among female claimants randomly assigned to female doctors relative to male doctors. Unlike the baseline specification, this specification focuses only on female claimants and thus does not partial out across-the-board differences in evaluations by male and female doctors.

4.2 Identifying Variation

Since the doctor evaluating a claimant is random conditional on the claimant’s county and the doctor’s credential, a claimant’s characteristics that are pre-determined prior to the initial designated doctor evaluation should not be associated with the designated doctor’s gender after controlling for the doctor’s credential and claimant’s county. An advantage of our setting is that we have rich baseline data on pre-determined claimant observable characteristics to verify that the assignment of providers appears random and orthog-

onal to baseline claimant characteristics.

We estimate Equation (1) replacing the dependent variable with baseline claimant characteristics. In this analysis, we examine the claimant's age as well as baseline claim characteristics that are determined before the designated doctor assignment, including measures that capture the severity of the injury: an indicator for whether the claim originated with an emergency department visit, medical spending on the first day of claim, total medical spending prior to the designated doctor exam, an indicator for whether the claimant received income replacement benefits prior to the designated doctor exam. We also examine two additional claim characteristics: the impairment type and the duration of time between the injury and the designated doctor exam.

Table 3 presents the resulting coefficient estimates for the key terms: the interaction of female doctor and female claimant (column 1), female doctor (column 4), and female claimant (column 7). The table also reports the associated standard errors and p-values. There are a few patterns to note. First, as expected, the assigned designated doctor's gender is uncorrelated with baseline worker and injury characteristics. We see this in Table 3 as the point estimates are small and statistically indistinguishable from zero for female doctor (column 4) and the interaction of female doctor and female claimant (column 1). This suggests the random assignment of designated doctors was implemented as required by state regulation. Second, the coefficient estimates for female claimant are often statistically distinguishable from zero. Female claimants are different than male claimants within the population of claimants receiving designated doctor exams, and these coefficient estimates capture the mean difference across female and male claimants in baseline observable characteristics. Many of these differences reflect differences between female and male workers—and workers' compensation claimants—more generally.²³ Throughout our analysis, we include a female claimant control to account for any systematic differences in the propensity of designated doctors to award further benefits to female claimants relative to male claimants.

5 Differences in Evaluations by Doctor Gender

5.1 Baseline Results from Randomized Evaluations

We now examine the relationship between the gender of the evaluating doctor and the main outcomes of interest: being assessed as having a continued disability (or equivalently, receiving any additional cash benefits) and the subsequent normalized cash benefits received. Table 4 displays estimated coefficients on the female doctor indicator variable, the female claimant indicator variable, and the interaction of the female doctor and female claimant indicator variables from Equations (1), (2), and (3).

The first three columns of Table 4 show OLS estimates from regressions with the dependent variable being an indicator variable equal to one if the designated doctor assesses the claimant as having a continued disability. Column 1 displays estimates from Equation (1). The coefficient of -0.031 on the female claimant variable indicates that female claimants are 3.1 percentage points less likely than male claimants to be assessed as having a continued disability when both are assessed by male doctors. This raw gender gap when male doctors evaluate claimants could be driven by a number of factors, as female and male claimants differ from one another for several reasons. While female and male claimants differ from each other, the random assignment of claimants to designated doctors means that claimant characteristics are orthogonal to characteristics of their assigned designated doctors. Thus, any systematic difference in disability assessments between male and female doctors is not the result of systematic differences in the claimants they evaluate.

²³See discussion of Table 1 in Section 3 for a comparison of female and male workers' compensation claimants, both in the universe of claimants and among the subset of claimants with independent medical evaluations.

The coefficient on the female doctor indicator variable in column 1 of Table 4 is the difference in the likelihood that male claimants are assessed as having a continued disability when examined by a female doctor rather than a male doctor. The small and statistically insignificant coefficient on the female doctor indicator variable of -0.0002 suggests that male and female doctors make similar disability assessments when evaluating male claimants. In other words, the gender of the assigned doctor does not impact whether male claimants are evaluated as disabled.

In contrast, there is a sizable impact of the gender of the assigned doctor on outcomes for female claimants. Compared to differences among their male patient counterparts, female patients randomly assigned a female doctor rather than a male doctor are 3.0 percentage points [95% C.I.: 1.0 to 5.0 percentage points] more likely to be evaluated as having a continued disability, or 5.0% of the mean among female patients evaluated by male doctors. These estimates are large in magnitude, with the 3.0 percentage point increase in the likelihood of being evaluated as disabled representing a near complete closure of the gender gap when male doctors are the evaluators.

Given that claimants are randomly assigned to doctors, the estimated coefficient on the female doctor and female claimant interaction term should be similar in models that include doctor fixed effects. Column 2 of Table 4 shows estimates from Equation (2), which includes doctor fixed effects, and verifies that the analysis yields similar coefficient estimates on the female doctor and female claimant interaction term when focusing on within-doctor differences in assessments for female and male claimants. Column 3 of Table 4 restricts the sample to include only female claimants and provides additional evidence that female claimants are more likely to be assessed as being impaired when seen by a female doctor rather than a male doctor.

While designated doctors have sole discretion over whether claimants receive any additional cash benefits after the designated doctor exam, employers, insurers, treating doctors, and claimants can all influence the final amount of cash benefits that claimants assessed as having a continued disability eventually receive. However, designated doctors can still influence benefit amounts, both through their initial assessments and because a claimant's original designated doctor is responsible for any future independent medical evaluations. Columns 4 through 6 of Table 4 consider how designated doctors' gender is related to the additional benefits claimants receive after the exam by showing estimates from Poisson regressions with normalized additional benefits after the exam as the dependent variable. Relative to differences among male patients, female patients randomly assigned a female doctor rather than a male doctor receive 8.5% more cash benefits on average—or \$478 dollars evaluated at the mean of \$5,627. This effect is statistically significant, with the 95% confidence interval spanning 2.6% to 14.4%. This effect is also sizable, with the estimated 8.5% increase in cash benefits representing a 61% closure of the gender gap in cash benefits observed when male doctors evaluate patients. There is no analogous gender-match effect for male patients. In fact, the point estimates suggest that male patients may receive more subsequent cash benefits when assigned a female doctor—3.1% more on average—though this effect is statistically indistinguishable from zero and is small relative to the 11.6% increase in mean subsequent benefits for female patients assigned a female doctor. Including doctor fixed effects in column 5 provides additional evidence that female claimants receive more benefits when assigned to female doctors than when assigned to male doctors. When we restrict the sample to females in column 6, the estimate indicates that female claimants receive 11.6% more in benefits when evaluated by a female doctor rather than a male doctor.

The estimates indicate that female claimants receive more workers' compensation benefits after being evaluated by a female designated doctor rather than a male designated doctor, compared to the analogous

differences in evaluations among male claimants. The random assignment of claimants to doctors means that these differential assessments are unlikely to be driven by either observed or unobserved differences in pre-determined characteristics of the claimants assigned to male and female doctors. As discussed above, we confirm the robustness of our basic findings when including doctor fixed effects or excluding male patients from the estimation data. In Appendix Table A2, we further probe the robustness of our findings to the inclusion of different combinations of fixed effects. The results are very similar regardless of the combination of fixed effects included. For instance, we obtain similar estimates when replacing separate county-credential and exam year fixed effects with county-credential-year fixed effects. This rules out alternative explanations related to the pool of designated doctors changing over time. We also obtain similar estimates in a specification including county-doctor-year fixed effects, in which the coefficient on the female doctor and female claimant interaction term is identified using differences in outcomes between male and female claimants who are assigned the same doctor from the same pool of potential designated doctors. Further, the estimates are similar in specifications with interactions between claimant demographic characteristics and time, which validates that the documented differences are not driven by unobserved factors related to female claimant evaluations that vary across time or space. We also obtain similar estimates when we replace the county-by-credential fixed effects with county-by-diagnosis fixed effects, an alternative approach to isolating conditional random assignment.²⁴ Finally, Appendix Table A2 also shows that we obtain similar estimates if we supplement the baseline specification with insurer-by-year fixed effects, which rules out changes in insurer practices across time as an explanation of the results.

In summary, the findings indicate that being evaluated by a female doctor (relative to a male doctor) improves outcomes for female claimants and reduces gender gaps in outcomes relative to those observed when evaluations are conducted by male doctors. The magnitudes of the estimates are large relative to observed gender gaps in both outcomes—the likelihood of being evaluated as disabled and total cash disability benefits received. Overall, the estimates indicate that the gender of the evaluating doctor is an important determinant of gender gaps in this setting, and the under-representation of females among doctors is likely a major explanation behind observed gender gaps in these outcomes. Additional analysis below illustrates these findings are similar when controlling for available baseline information about claimants. Section 8 further explores the implications of our findings for gender gaps through associated policy counterfactuals.

5.2 Mechanisms and Additional Evidence from Randomized Evaluations

There are several reasons why gender match between doctors and patients could matter for medical evaluations. Differences in evaluations across male and female doctors may reflect provider discrimination against patients of the opposite gender or favoritism towards patients of the same gender. For instance, doctor-patient gender-match effects could arise if doctors disproportionately discount medical conditions of patients of the opposite gender or disproportionately empathize with patients of their own gender. Further, differences in evaluations could arise if patients behave differently when matched with providers of the same gender, perhaps communicating more or different information that may affect their evaluation. We note that it is not possible to conceptually (or empirically) distinguish between these mechanisms in this setting. For example, patients matched with own-gender doctors may communicate more or different information during an exam if doctors differentially empathize with same-gender patients and ask more

²⁴See Section 3 for a detailed description of the designated doctor assignment mechanism and alternative approaches to empirically isolate conditional random assignment.

questions when evaluating these patients.²⁵ Nevertheless, it is important to emphasize that doctors are ultimately responsible for outcomes in the medical examinations we analyze after gathering information from a physical exam, patients' prior medical records, and patients' own accounts of their condition and work limitations. Thus, differences in evaluation outcomes in our setting are driven by differences in doctor decisions, regardless of whether the reasons for doctors' differential decisions stem from differences in doctor behavior, patient behavior, or a combination of doctor and patient behavior.

More generally, health care decisions typically involve complex interactions between patients and doctors, with health care interactions spanning a spectrum between doctor and patient agency. For example, while patients are ultimately responsible for deciding whether to take up recommended preventive care, doctors are responsible for performing the medical evaluations we examine. The relative agency in most health care interactions lies between these extremes—with the doctor first evaluating the patient's condition, the doctor then recommending treatment based on this evaluation, and the patient deciding whether to follow the doctor's treatment recommendations. Our work provides the first evidence on the role of demographic concordance between doctors and patients in doctors' evaluations of patients' medical conditions—the critical initial step in most health care interactions (and in disability determinations) and an aspect of health care interactions where doctor agency is central. In this way, our work complements prior work by Alsan, Garrick, and Graziani (2019), which investigates the role of doctors' race in health care interactions of black men in a setting where patient agency is central—patient decisions to take up effective and broadly recommended preventive care. The mechanisms behind the importance of demographic concordance in health care settings may differ across different types of health care interactions, depending on the relative agency of doctors and patients. While Alsan, Garrick, and Graziani (2019) argue that the most plausible explanation for why black men increase take up of preventive care when paired with a black doctor rather than a white doctor is by improving communication between patients and providers, the plausible mechanisms behind the importance of gender concordance on outcomes in our setting may be different and more complex given the central role of doctor agency in the medical evaluation outcomes we examine.

In our setting, we find that gender match between doctors and patients increases evaluated disability and subsequent cash benefits for female patients, while gender match does not matter for the evaluation of male patients. This suggests that mechanisms behind the importance of gender match—such as reduced discrimination, increased favoritism, or improved communication—do not operate symmetrically for men and women. Below, we explore potential mechanisms further through additional analysis within the setting of these randomized evaluations. This supplemental analysis reveals three main findings: (i) there appears to be something fundamental about gender match that affects evaluations of women (as opposed to being driven by correlated patient or provider characteristics), (ii) the gender-match effect is near universal, affecting many subgroups of women, and (iii) female doctors similarly evaluate men and women with similar observable characteristics while male doctors evaluate women as less disabled than men with similar observables. To complement our analysis with the administrative data, Section 7 presents evidence from a complementary survey we conducted to further explore plausible mechanisms. Overall, the survey evidence supports the plausibility of the main finding: that doctor gender is an important factor in the evaluation of female patients more so than in the evaluation of male patients. Further, the survey evidence

²⁵Because both providers and patients may contribute to gender differences in evaluations and we do not have an objective measure of disability that is not influenced by providers, we do not attribute all of the observed differences in evaluations to discrimination—e.g., providers discriminating against opposite-gender patients or empathizing with same-gender patients—and thus it is not conceptually or empirically feasible to distinguish between statistical or taste based discrimination in this setting.

suggests differences in doctor behavior likely contribute to the main findings and highlights several possible ways male doctors—relative to their female counterparts—may discount the concerns and symptoms of female patients.

Other Provider and Patient Characteristics It is possible that the gender-match effect we document arises because of some other patient or doctor characteristic that is correlated with patient or doctor gender rather than mechanisms inherently linked to gender. We consider this possibility by conducting two sets of supplemental analysis. First, we assess whether female and male doctors treat patients differently by observed patient characteristics other than patient gender. Table 5 replicates the baseline analysis but replaces the female claimant indicator variable with other claimant characteristics. The various characteristics considered are indicated as column headings. We consider demographic characteristics including pre-injury income, age, and marital status; injury characteristics such as whether the injury is associated with a sprain or muscle issue, a dangerous industry, or higher than median predicted cash benefits; and characteristics of the patient’s prior medical and claim experiences such as whether the patient had previously seen a female doctor, had requested a designated doctor exam, or had higher than median first day medical spending.²⁶ Row 1 of the estimates displays coefficients on the interaction of the indicated characteristic and the female doctor indicator variable. Row 2 reports coefficients on the female doctor indicator variable. Row 3 reports coefficients on the indicated characteristic. For all characteristics considered in Table 5, the interaction of female doctor and the additional characteristic is statistically indistinguishable from zero. These results suggest that male and female doctors do not differ in their assessments based on claimant characteristics other than claimant gender.

Second, we examine whether doctor characteristics other than doctors’ gender can explain the differential assessments of female claimants by male and female doctors. To do this, we first regress the indicator variable for doctors assessing claimants as having continued disability on injury year fixed effects, exam year fixed effects, doctor fixed effects, and separate doctorXfemale-claimant fixed effects for each of the 917 doctors with at least ten exams in the sample. We then regress the estimated doctorXfemale-claimant fixed effects—a doctor-specific measure of disability assessment generosity toward female patients relative to male patients—on doctor characteristics in doctor-level regressions weighted by the number of exams doctors performed. These regressions inform us about the relationship between observed doctor characteristics and doctors’ differential assessments for female claimants. Table 6 displays these estimates. Column 1 shows the coefficient on a female doctor indicator variable from a regression with no additional controls and indicates that a doctor’s differential generosity toward female patients is higher on average for female doctors than for male doctors. Columns 2 through 7 supplement the specification in column 1 with an indicator variable for the designated doctor being an MD or DO rather than a DC, an indicator for the doctor’s MD or DO degree being from a medical school in the top 25 of all U.S. medical schools according to U.S. News and World Report, indicators for the doctor’s specialty, doctor experience, and an indicator for the doctor being born in Texas. Of all the doctor characteristics considered, only doctors’ gender appears to be associated with differential generosity in assessments of female claimants, and including additional doctor characteristics does not result in meaningful changes to the coefficient on the female doctor indicator

²⁶We classify claimants working in agriculture, mining, construction, manufacturing, transportation, or warehousing as being in dangerous industries. To create the measure of predicted cash benefits being above the median, we first fit a lasso model of normalized cash benefits where benefits are measured through one year after the designated doctor exam. We then create an indicator variable that is equal to one if claimants’ predicted cash benefits based on this lasso model are above the median value. For the lasso model, we include indicator variables for ten-year age bins, wage deciles, day of the week of first medical treatment, industry, and injury type. We also include the cost of first-day medical treatment and indicator variables for gender, for first treatment occurring in the emergency department, and for marital status.

variable.

In addition to the possibility that fixed characteristics of doctors could relate to evaluations of females relative to males, doctors' experience performing these exams also has the potential to influence evaluations. For example, doctors could potentially become more consistent in evaluating female claimants after having evaluated more female claimants. In Appendix Table A1, we show results from specifications that supplement the baseline specification of Equation (1) with controls for a doctor's recent experience with independent medical exams—the number of independent medical exams that the designated doctor performed in the prior year and the share of those exams that were for females—and interactions between these experience measures and an indicator for female claimant. We note that conditional on the included controls, the share of a doctor's exams for female claimants is random. In Appendix Table A1, we see that the key coefficient capturing the gender-match effect—the coefficient on the female doctor by female claimant interaction term—is stable when adding these additional controls, and there is no evidence that recent experience with exams impact male-female differences in evaluated disability. This evidence suggests that the gender-match effect is not explained by differences in doctors' experience with independent medical exams and that doctors' experience with these exams does not predict gender differences in evaluation outcomes.

Taken together, the evidence in Tables 5, Table 6, and Appendix Table A1 suggests that there is something fundamental or essential about gender—separable from other observed patient and doctor characteristics—that is driving the gender-match effect we document.

Heterogeneity Analyses Heterogeneity analyses may shed light on possible mechanisms. For instance, if gender match is only relevant among particular subgroups of female patients, this may point to potential mechanisms behind the importance of gender match. We explore whether the gender-match effect is more concentrated among some subgroups of patients than among others, based on demographic characteristics, injury characteristics, and prior medical experiences. Figure 1 considers how the effect of female claimants being assigned to female doctors differs by baseline claimant characteristics by showing estimated coefficients on the interaction of female doctor and female claimant from the baseline specification estimated using data from the indicated subgroups. Panel A displays the estimates for an indicator of any additional cash benefits after the exam (or equivalently, being evaluated as disabled at the time of exam), while Panel B displays the estimates for normalized additional cash benefits. Capped horizontal bars indicate the associated 95% confidence intervals.

The first set of results in Figure 1 displays separate coefficients on the interaction of the female claimant and female doctor indicator variables for claimants with different demographic characteristics including income, age, and marital status. The second set of results in Figure 1 displays separate female match coefficients based on characteristics of claimants' injury and work including: whether the injury was a sprain or muscle issue, whether the injury took place in a dangerous industry, and whether the injury is above or below the median in terms of predicted cash benefits—a proxy for the duration of injury-related disability. The final set of results shown in Figure 1 considers possible heterogeneity based on prior medical and claim experience. In this set of results, we examine heterogeneity by whether the claimant previously had a female treating doctor, whether the claimant had requested a designated doctor exam, and whether the claim had above or below median medical spending on the first day of the claim—a proxy for severity of the injury.

Examining Figure 1, it is striking that the gender-match effect estimates are positive within all subgroups considered and statistically distinguishable from zero for many subgroups. These findings suggest that patient-doctor gender match has near universal impacts on outcomes for female patients. Though the

estimates across subgroups are not statistically distinguishable from one another, the pattern in the point estimates suggests that the female gender-match effect may be more pronounced among claimants who have lower earnings, had a female treating doctor, have more severe injuries, did not initiate a designated doctor exam, and work in less dangerous industries.

Impact on Gender Gaps Conditional on Observables Because there is no objective measure of the appropriate level of evaluated disability or efficient level of cash benefits conditional on evaluated disability, we cannot determine whether female doctors are too generous or male doctors are too strict in their evaluations of female claimants.²⁷ While we do not have a comprehensive objective disability measure to benchmark evaluations against, we can leverage rich baseline data on claimants to investigate whether patient-doctor gender match ameliorates or magnifies differences in the evaluation of male and female patients conditional on observables. Specifically, we re-estimate our main econometric model adding successively more controls for claimant and injury characteristics: insurer fixed effects, the day of the week on which the injury occurred, age indicators, diagnoses indicators (indicators for three-digit ICD-9 code), indicator for whether the first injury-related treatment occurred in the emergency department, and the inverse hyperbolic sine of medical spending on the first treatment date.²⁸ Before discussing this evidence, we note two important caveats. First, though we observe rich baseline data on claimants, male and female claimants may still differ along unobserved dimensions related to benefit receipt. Second, while this analysis sheds light on differences in evaluations of female claimants relative to observably similar male claimants, we cannot make normative statements about how the benefit amounts resulting from these evaluations compare to optimal benefit levels. In addition to needing an objective benchmark of disability that does not rely on a doctor's evaluation, we would also need a framework for determining the optimal benefit payments for different levels of objective disability.

The results are shown in Tables 7 and 8. There are several patterns in these tables worth noting. First, including richer controls has little impact on the magnitude or precision of the coefficient on the interaction of female doctor and female claimant, confirming the robustness of our main findings regarding the gender-match effect. Second, the female claimant coefficient is consistently negative, indicating that there are significant gender gaps conditional on observables across female and male patients when they are evaluated by male doctors. When controlling for all available covariates (in column 7), the magnitudes indicate that female patients evaluated by male doctors are 2.6 percentage points less likely to receive benefits and receive 10.4% fewer benefits than male patients evaluated by male doctors. Third, the coefficients on the interaction term and the female claimant term are of nearly the same magnitude and opposite sign, indicating that having female doctors evaluate claimants shrinks gender gaps substantially—to the point of

²⁷Further, note that our analysis does not speak to whether benefits may optimally vary by individual characteristics including gender. Even among workers with the exact same disability and prior earnings, optimal benefit levels may vary across males and females, as the social planner may consider other factors in optimal benefit design—such as the magnitude of the consumption drop experienced by workers upon workplace injury or the degree of behavioral responses to benefit generosity. Note, however, that doctors in this setting (and gatekeepers more generally in other disability-related settings) are tasked with objectively evaluating disability rather than considering outside factors. Though beyond the scope of this paper, how individual and injury characteristics may influence optimal benefit design is an important area for future work. Further, it is important to note that from an injured workers' perspective, being evaluated as more disabled improves outcomes for that worker through increased cash benefits. In this way, our estimates indicate that gender match between doctors and claimants improves outcomes for women and shrinks gender gaps in disability benefit receipt, even if broader efficiency implications are more difficult to assess.

²⁸There are 466 distinct insurers represented in the disputed claims sample. Because certified self-insured employers have their own insurer codes, insurer fixed effects are equivalent to controlling for an employer fixed effect among claimants working for some large employers. Claims' ICD-9 codes are identified using medical bills from the first day of medical treatment. For bills that identify ICD-10 codes, we convert ICD-10 codes to ICD-9 codes using a crosswalk from the Centers for Medicare & Medicaid Services so that we have consistent definitions of injuries over time. We create a separate indicator variable for the 2,358 claims without a valid ICD-9 code from the first medical treatment, but the results are similar if we instead drop these claims or impute codes using information from subsequent treatment.

almost eliminating these gaps. Relative to evaluations performed by male doctors, evaluations performed by female doctors eliminate the gender gap in evaluated disability and shrink the gender gap by 73% for total additional disability benefits received. These results indicate that gender of evaluating doctors is a key determinant of gender gaps conditional on observables in this setting.

In specifications controlling for all available baseline observable information about the patient, we see female doctors evaluate female and male patients as similarly disabled while male doctors evaluate female patients as less disabled than male patients. While this evidence is only suggestive, these results are consistent with male doctors evaluating female patients against a stricter standard than male patients and female doctors applying similar standards to male and female patients.

6 Broader Evidence: Gender Gap and Gender Homophily

Our primary analysis estimating the impact of gender-match between doctors and patients on outcomes focuses on disputed claims with randomly assigned evaluating doctors. While disputes over workers' compensation cash benefits are common—39% of all workers' compensation claims with cash benefits are disputed—and the resolution of disputes is interesting in its own right, we may wonder whether there are gender disparities in disability evaluations and cash benefit receipt more broadly among workers' compensation claims. An advantage of our setting is that we have administrative data on the universe of claims—all workers' compensation claims in Texas from 2013 to 2017. Next, we use data from the entire workers' compensation insurance system to provide broader evidence on gender disparities in workers' compensation insurance cash benefit receipt and gender homophily in the selection of doctors when patients have choice.

Gender Gap in Benefit Receipt We use data from all workers' compensation claims to investigate whether gender disparities in cash benefit receipt are observed more broadly among workers' compensation claimants. While all injured workers qualify for workers' compensation medical benefits, cash benefit eligibility is limited to workers with a qualifying disability and this eligibility is determined by a worker's treating doctor (or in the case of disputed claims, the randomly assigned designated doctor). As discussed in Section 3, workers are eligible for cash benefits if the doctor determines that the worker has either: (i) a temporary impairment that leaves him/her unable to work for more than seven days, or (ii) a partial permanent impairment. Approximately 22% of all workers' compensation claimants qualify for cash benefits, while the remainder have "medical only" claims that only involve reimbursement for injury-related medical expenditures.

We examine how the propensity to receive any cash benefits varies across male and female claimants. Specifically, we estimate the following equation:

$$I(\text{cash benefits}_i > 0) = \beta \text{female_claimant}_i + \Theta X_i + \epsilon_i, \quad (4)$$

where i denotes claimant, $I(\text{cash benefits}_i > 0)$ indicates if the claimant receives any cash benefits (temporary income benefits or permanent impairment benefits), and female_claimant_i indicates the claimant is female. We illustrate how the estimated gender gap varies with the included additional controls, X_i .

Table 9 displays the estimates. Column 1 displays the estimates with no additional controls. Columns 2 through 6 display the estimates with progressively more controls. These additional controls include information about the claim (insurer), about the claimant (county of residence and age), and about the injury including injury timing (month and day-of-the-week), type (injury type), and severity (an indicator for first medical treatment taking place at an emergency department).

According to the estimates with no controls (from column 1), on average female claimants are 5.5 percentage points less likely to receive benefits than male claimants, which represents a 23.3% reduction relative to the overall mean rate of cash benefit receipt among male claimants (23.6%). When controlling for claim characteristics, we see the gender difference in benefit receipt remains substantial, and the estimates are precise in all specifications with the standard error never exceeding 0.1 percentage points. Across the range of controls considered, the estimates indicate that females are between 3.3 to 4.0 percentage points less likely to receive cash benefits than males. The estimate from a specification including the full set of claimant and injury controls indicates that female claimants are 3.9 percentage points less likely to receive cash benefits, or 16.5% lower than the mean rate of benefit receipt among male claimants.

The main takeaway from this analysis is that female claimants receive cash benefits at a substantially lower rate than male claimants with similar observables. This evidence points to the broader importance of gender disparities in cash benefit receipt among workers' compensation claimants, beyond the subset of claims with randomized evaluations.

Gender Homophily in Choice of Providers More generally, workers' compensation claimants have the ability to select their own "treating doctor," the doctor who typically oversees their medical care and their eligibility for disability benefits. Our main analysis focuses on disputed claims where a claimant or a claimant's insurer disagrees with the treating doctor's evaluation, and in such cases the final benefit determination is made by a randomly assigned designated doctor. Using this random assignment, our main analysis illustrates that gender match increases evaluated disability and subsequent cash benefit receipt for women. If women understand that they typically fare better when evaluated by a female doctor rather than a male doctor, we might imagine women would express a preference for female doctors relative to male doctors and select female doctors more frequently when given the choice.

We investigate whether patient-provider matches exhibit relative gender homophily. Given the differing characteristics of female and male doctors (e.g., specialty, availability), we focus on measuring relative gender homophily—do female patients select female treating doctors more often than male patients select female treating doctors? Specifically, we estimate the following equation:

$$I(\text{Chosen Treating Doctor is Female})_{im} = \gamma \text{female_claimant}_i + \sigma X_{im} + \delta_m + \mu_{im}, \quad (5)$$

where i denotes claimant and m denotes the claimant's medical market. In this equation, $I(\text{Chosen Treating Doctor is Female})_{im}$ indicates that the chosen treating doctor is female and female_claimant_i indicates the claimant is female. Most of this analysis focuses on comparing claimants within the same medical market, to focus on differences in selections of male and female patients facing the same available pool of doctors. To do this, we include patient medical market fixed effects, δ_m , where we define medical markets following the Dartmouth Atlas' definition of Hospital Service Areas (HSAs). We also include additional controls in some specifications (X_{im}), including insurer, injury month-year, injury diagnoses codes, claimant age, and an interaction between injury month-year, diagnoses, and patient medical market. The sample includes all claims for workers in Texas with an identifiable treating doctor.²⁹

²⁹We identify a claimant's treating doctor as the doctor who submits a workers' compensation report describing a claimant's ability to work. If a claimant does not have a bill that explicitly states that a work status exam was performed, we identify the claimant's treating doctor as the doctor who bills for case management services. If multiple doctors bill for case management services for a claim, we identify the treating doctor as the first doctor to bill for case management services. We can identify the treating doctor and the treating doctor's gender for about half of claims in our sample using this approach. For the other half of claims, either no case management services were billed, which often happens for medical-only claims (i.e., claims without cash benefits), or the case management services were billed to a health care organization rather than to a specific provider (in which case, we cannot determine the gender of the treating doctor). We obtain similar results if we instead adopt a treating doctor definition based on first office visit,

Table 10 displays the estimates. Column 1 displays the estimates with no controls. The remaining columns include patient medical market fixed effects and progressively add additional controls. Without controlling for patient medical market, we see that female patients are 2.3 percentage points more likely to select a female treating doctor, which is an 8.3% increase relative to the share of male patients selecting a female treating doctor (27.6%). This relationship is precisely estimated, with a standard error no more than 0.2 percentage points across all specifications considered. Once we control for patient medical market, the effect is roughly 78% of the size indicating that some of the female coefficient estimate in the specification without patient medical market fixed effects reflects a correlation between patient and doctor characteristics across markets: markets with a greater share of female patients (female injured workers) are markets with a greater share of female doctors. Among female and male patients within the same medical market, females are 1.8 percentage points more likely to select a female doctor than their male counterparts, or 6.5%. After controlling for patient medical market, we see that the estimated difference in the rate that females and males select female doctors is very stable as we add more controls. In particular, we obtain a similar estimate—that females are 1.6 percentage points (5.8%) more likely to select a female doctor—in the most saturated specification which controls for patient age and insurer fixed effects and compares female and male patients with same types of injuries, within the same medical market, at the same point in time.

We note that this analysis identifies relative gender homophily, and the resulting patterns could be driven by preferences of either men or women to see own-gender providers. In Section 7.2, we present survey evidence that suggests that these patterns may be largely driven by women’s preferences for own-gender providers. Specifically, the survey evidence suggests that women—compared to men—have stronger preferences for own-gender providers, more often seek care from own-gender providers, and report being willing to pay more to see an own-gender provider.

7 Survey Evidence

Next, we describe the results of a survey we conducted that had two aims. Our first aim is to assess differences in male and female patients’ experiences in the broader health care system and to learn about how and why doctors’ gender is associated with differential health care experiences for male and female patients. The questions speak to possible mechanisms behind our main finding—that gender match between doctors and patients increases evaluated disability for female patients with no analogous effect for male patients. A second aim of the survey is to understand whether patients exhibit preferences for own-gender providers and whether the intensity of these preferences varies across men and women. The questions within this part of the survey can help us understand the generalizability of the gender homophily patterns documented in the prior section and assess potential explanations for these patterns.

We conducted a survey of 1,519 adults between 30 to 64 years of age. The survey was created and conducted using the Qualtrics platform. Qualtrics aims for its panel to include a diverse set of respondents across age, gender, ethnicity, household income, and census region. To focus on a sample that overlaps with individuals at risk of a workplace injury, we restricted the pool of eligible respondents to people who have worked in the last 12 months, and we targeted an age distribution of respondents to roughly match the age distribution of workers in the American Community Survey (ACS). We limited the survey to individuals who identify as either male or female, and we had roughly equal numbers of respondents in each group, with 755 male respondents and 764 female respondents. The survey was conducted in May of 2021.

which likely misidentifies treating doctors more often than our baseline approach does but leads to fewer claimants having missing treating doctor gender.

The survey asks respondents about their past experiences with health care providers and about how a doctor’s gender influences their expectations about various measures of quality of care based on their past experiences with the health care system. The survey also explicitly asks people about their preferences over doctors’ gender and about the importance of doctor gender and other doctor attributes when selecting a doctor. In addition, the survey collects basic information on respondents’ amount of health care used, health insurance coverage, employment, and demographics. We took several standard steps to ensure high-quality responses, including dropping respondents who completed the survey in less than one half of the median time to completion and including a screening question that excludes respondents who do not commit to providing thoughtful and honest answers.

Appendix Table A3 reports respondent characteristics summarizing responses to the survey questions about the amount of health care used, health insurance coverage, employment, and demographics. In addition to summarizing respondent characteristics by gender, the table also shows analogous summary statistics by gender for workers in the 2019 ACS and for workers’ compensation claimants from the Texas administrative data when overlapping information is available. There are a few patterns worth noting. Broadly speaking, the survey respondents and workers more generally in the ACS appear similar on basic demographic characteristics, the propensity to have health insurance, industry composition, and geographic representation. The age distribution of respondents overall looks similar to workers more generally and workers’ compensation claimants, though there are some differences in the age distribution of survey respondents by gender with an over-representation of female respondents 30 to 39 years old and male respondents 40 to 49 years old. While survey respondents report a similar range of educational backgrounds and earnings as workers more generally, on average survey respondents are somewhat more educated and earn slightly more than workers overall. Because there are differences in observable characteristics across male and female respondents, we report survey responses in two ways—raw means by gender and regression-adjusted differences by gender which account for variation in observable characteristics.

Tables 11 and 12 summarize responses to the questions about experiences with the health care system and preferences over providers, respectively. These two tables are laid out similarly, where columns 1 and 2 report mean responses separately for female and male respondents. Column 3 reports the difference in mean outcomes between female and male respondents, and column 4 reports the p-value for the difference. To control for the influence of underlying differences in health, employment, or demographics on female-male differences, we also estimate the following model:

$$y_i = \gamma female_i + \sigma X_i + \mu_i, \tag{6}$$

where y is the indicated response, $female$ indicates a female respondent, and X is the set of characteristics shown in Appendix Table A3. The coefficient on $female$ is the regression-adjusted difference in female-male responses and is reported in column 5 of Tables 11 and 12. The standard error and p-value on the $female$ coefficient are displayed in columns 6 and 7, respectively.

7.1 Gender Differences in Experiences with the Health Care System

The survey begins with questions about past experiences with health care and respondent expectations about health care interactions with male and female doctors informed by their past experiences. Table 11 summarizes the responses to these questions. Overall, the responses to these questions highlight a few broad patterns. First, women are more likely than men to report a number of negative health care experiences have ever happened to them. Second, women are much more likely than men to report that they

expect positive interactions are more likely with own-gender doctors than with opposite gender doctors. Collectively, this survey evidence supports the main finding of the paper—that gender match between patients and doctors impacts medical evaluations for women—and points to several plausible mechanisms behind this finding.

The first set of questions asks respondents whether they have ever experienced various negative interactions with a health care provider. The main takeaway from this series of questions is that women are much more likely than men to report a negative experience with the health care system. For example, 39.3% of women and 34.0% of men report having an experience where a doctor talked down to them or didn't treat them with dignity or respect (a greater than 5 percentage point difference). Women are 9.4 percentage points more likely to report encountering a doctor who didn't understand or relate to their experiences and concerns and 4.2 percentage points more likely to report a doctor made them feel uncomfortable discussing their concerns. Slightly more women than men reported a doctor didn't believe they were telling the truth about their symptoms and reported a doctor assumed something about them without asking. Generally, the documented gender differences are even larger when accounting for differences in observable characteristics between male and female respondents: an 8.7 percentage point difference in reporting a doctor talked down to them (p-value 0.001), a 10.4 percentage point difference in reporting a doctor did not understand or relate to their concerns (p-value less than 0.001), a 4.5 percentage point difference in reporting a doctor didn't believe they were telling the truth about symptoms (p-value 0.07), a 5.9 percentage point difference in reporting a doctor made them feel uncomfortable discussing their concerns (p-value 0.01), a 5.0 percentage point difference in reporting a doctor assumed something without asking (p-value 0.05).

As a summary measure, we also report the share of respondents who experienced at least one of these negative experiences. Women are more likely to report they have experienced any of the negative experiences asked about in the survey, with a gap of 5.8 percentage points in the raw means and 10.3 percentage points after controlling for differences in respondent characteristics. While most of these negative experiences were more prevalent among women than men, there were two exceptions: women and men were equally likely to report that a doctor refused a test or treatment they thought they needed and men were more likely to report that a doctor had refused to prescribe pain medication they thought they needed. We note this latter difference in responses could potentially stem from underlying differences in the likelihood of needing (or wanting) pain medication, though the difference remains statistically significant even after accounting for the controls which include information on health care use and an indicator for having experienced chronic pain.

In the next set of questions, the survey asks respondents about how doctors' gender influences their expectations about the likelihood of having a positive health care interaction along various quality measures based on their past experiences getting health care. Respondents can answer that they would expect a better experience with a male doctor or a female doctor or that doctors' gender is not informative about the likelihood of having a positive interaction.³⁰ Figure 2 (and Table 11) shows the answers two ways. First, we show the share of respondents saying a doctor of their own gender would be more likely to have the positive trait (Figure 2 Panel A). Second, we show the share of respondents who indicated a doctor of their own gender is at least as likely to have the positive trait—or equivalently, the share of respondents who did not indicate that a doctor of the opposite gender is more likely to have the positive trait (Figure 2 Panel B).

Relative to male respondents, a higher share of female respondents believe that gender concordance

³⁰To ensure that the ordering of options did not bias the responses, we randomized at the respondent level which choice—a male doctor or a female doctor—was listed first in these questions.

with their doctor would be more likely to result in a positive health care interaction than seeing a doctor of the opposite gender. Compared to male respondents, women are more likely to answer that a doctor of their own gender is more likely to: treat them with dignity and respect (34.3% vs. 19.3%), understand or relate to their experiences and concerns (51.7% vs. 29.5%), believe they are telling the truth about their symptoms (30.4% vs. 17.5%), provide needed testing and treatments (24.9% vs. 17.6%), make them feel comfortable discussing their concerns (41.6% vs. 24.0%), and ask appropriate questions instead of making assumptions (32.6% vs. 17.7%). These differences in responses across male and female respondents are large in magnitude and precisely estimated allowing us to reject the null hypothesis of no difference with a high degree of statistical confidence (the p-value never exceeds 0.001). Further, we note these differences are if anything larger when controlling for respondent characteristics. Overall, these responses indicate that females are more likely than males to report they expect more positive interactions with doctors of their own gender.

A natural follow-on question is whether these differences are also reflected in respondents' expectations of experiences with doctors of the opposite gender. Because the survey offered three answer options—male doctor, female doctor, or male and female doctors are equally likely—we can separately examine the share of respondents who expect a doctor of their own gender to be at least as likely to provide the stated positive interaction (or equivalently, the share of respondents who did not expect that a doctor of the opposite gender would be more likely to provide the stated positive interaction). We see the same patterns in expectations as above. Women are much more likely than men to report that doctors of their own gender are at least as likely to provide the stated positive interaction: treat them with dignity and respect (93.1% vs. 82.0%), understand or relate to their experiences and concerns (93.2% vs. 82.6%), believe they are telling the truth about their symptoms (92.8% vs. 81.6%), provide needed testing and treatments (93.2% vs. 85.3%), make them feel comfortable discussing their concerns (93.2% vs. 83.7%), and ask appropriate questions instead of making assumptions (93.3% vs. 81.3%). These gender differences are large, distinguishable from zero with p-values always less than 0.001, and robust to controlling for other respondent characteristics.

The final two positive attributes asked about within this set of questions pertain to which doctor is more likely to be available near you and to be most qualified. There is no meaningful difference between men and women in the share that respond that an own-gender doctor is more likely, or at least as likely, to be available nearby. A similar share of men and women report that an own-gender doctor is more likely to be the most qualified. However, a much larger share of women than men report that an own-gender doctor is at least as likely to be the most qualified (95.8% vs. 89.0%), amounting to a 6.8 percentage point gap with a p-value less than 0.001. This difference remains large and statistically distinguishable from zero when controlling for respondent characteristics.

7.2 Gender Differences in Preferences over Providers and Homophily

The next set of questions collects information about the propensity of respondents to select male and female doctors and about preferences regarding doctor gender. Table 12 summarizes responses to these questions. The responses collectively suggest there is homophily in patient-provider interactions, there is some preference for doctors of one's own gender, and this preference for gender concordance is stronger among women than men. It is perhaps not surprising that women more often prefer to see own-gender doctors given the evidence above on gender differences in expectations of positive experiences with own-gender versus opposite-gender doctors. Further, these findings help us interpret the homophily evidence in the prior section. While relative gender homophily can be driven by preferences of patients of either gender

for own-gender providers, this survey evidence suggests that women in particular prefer to see—and seek out care more often—from own-gender providers.

Respondents were asked whether they had ever received care from a male doctor and whether they had ever received care from a female doctor. There are notable differences by respondent gender, with 90.1% of women and 96.8% of men reporting they have received care from a male doctor while 94.1% of women and 76.7% of men report they have received care from a female doctor. These differences are statistically distinguishable from zero and robust to controlling for respondent characteristics. These patterns suggest there may be gender homophily in patient-provider matches more generally, pointing to the broader relevance of the homophily evidence presented in the prior section.

The next several questions ask respondents about their preferences related to doctor gender. The survey asks whether respondents prefer to see a doctor who is female, prefer to see a doctor who is male, or does it not make much difference. Women were more likely to express a preference for an own-gender doctor, with 46.3% of women preferring a female doctor while only 29.0% of men prefer a male doctor. The results also suggest that women are much less likely to report a preference for an opposite-gender doctor, with only 5.8% of women preferring a male doctor while 13.6% of men prefer to see a female doctor. The share of respondents expressing that doctor gender doesn't make much difference was smaller among women than men—47.9% among female respondents and 57.4% among male respondents. These differences are all statistically different from zero with p-values always less than 0.001 and are similar when controlling for respondent characteristics.

To better understand how the intensity of preferences over doctor gender varies across men and women, the next set of questions asks respondents to make a series of hypothetical choices over health care options. Specifically, respondents were told to suppose they needed to see a doctor for medical evaluation and treatment. They were then presented with two doctors who were identical in all other ways except for the listed characteristics, where the listed characteristics described the doctor's gender and the out-of-pocket cost for the visit. The respondents are presented with one of four choice sets:

- Female doctor, \$30 out-of-pocket cost or Male doctor, \$5 out-of-pocket cost
- Female doctor, \$10 out-of-pocket cost or Male doctor, \$5 out-of-pocket cost
- Female doctor, \$5 out-of-pocket cost or Male doctor, \$10 out-of-pocket cost
- Female doctor, \$5 out-of-pocket cost or Male doctor, \$30 out-of-pocket cost.

The initial choice set presented to each respondent is randomly assigned among those above. For completeness, we then ask sequential follow-up questions—moving either upward or downward from the initial choice set among the choice sets listed above—based on the respondent's answer to prior questions to discern the maximum the respondent is willing to pay for a doctor of a particular gender in terms of out-of-pocket cost differentials. Thus, each respondent is asked at most four hypothetical choice questions, though the median respondent was only asked 2 questions (mean 2.5 questions). To ensure that primacy bias did not lead to a higher percentage of respondents choosing either the male or female doctor option, we randomly assigned which option—the male doctor or the female doctor option—appeared first in each of the four choice sets above. The ordering of options was randomized in both the initially assigned hypothetical choice question and all follow-up hypothetical choice questions.

Figure 3 plots the share of respondents selecting an own-gender doctor by respondent gender. Panel A displays estimates based on only the initial randomly assigned choice set, while Panel B displays the esti-

mates based on the full set of hypothetical choice questions asked of respondents.³¹ Table 12 displays the underlying estimates. At every co-pay differential, we see that women are at least as likely as men—and typically strictly more likely than men—to select an own-gender doctor. Let us focus first on the results in Panel A based on the initial hypothetical choice question. Based on the responses, 48.5% of women are willing to pay an additional \$5 copay to see an own-gender provider, compared to only 29.3% of men. This 19.2 percentage point difference is statistically distinguishable from zero with a p-value less than 0.001. When it is cheaper to see an own-gender provider than to see an opposite-gender provider, nearly all women select to see an own-gender provider: 92.9% at an out-of-pocket cost differential of -\$5 and 98.8% at an out-of-pocket cost differential of -\$25. In contrast, a lower share of men would select to see an own-gender provider when a visit to an own-gender provider costs less: 83.3% at an out-of-pocket cost differential of -\$5 and 85.4% at an out-of-pocket cost differential of -\$25. These gender gaps are statistically distinguishable from zero and robust to controlling for respondent characteristics. We see similar patterns—though more precisely estimated—if we rely on responses to the full set of hypothetical choice questions displayed in Panel B. Overall, these results show a stark asymmetry between men and women in the intensity of preferences for own-gender doctors. Women appear to consistently place a higher value on seeing an own-gender provider than their male counterparts do—in terms of additional out-of-pocket costs they are willing to bear.

The survey also asks respondents to rate the importance of various doctor attributes—including a doctor’s gender—when they select a health care provider. Table 12 reports the share of respondents that rated each attribute as at least moderately important.³² Doctor gender was rated as at least moderately important by 41.4% of women compared to 33.6% of men—a more than 7 percentage point gap (p-value 0.002). This gap is larger—11.2 percentage points (p-value less than 0.001)—when controlling for respondent characteristics. This difference stands in contrast to the similarity of men’s and women’s responses for other doctor attributes, such as out-of-pocket cost for a visit, travel time to get to doctor’s office, wait time at the doctor’s office, and doctor’s age. Aside from doctor’s gender, we see gender differences in the stated importance of doctor reviews, with 84.7% of women indicating doctor reviews are at least moderately important compared to 80.7% of men, a 4.0 percentage point difference (p-value 0.038). Women placing more weight on doctor reviews could be in response to their greater propensity to have negative experiences with health care providers in the past.

7.3 Survey: Summing up

The survey evidence suggests that women—relative to men—more often have their concerns and symptoms discounted by doctors. For example, the survey responses indicate women were more likely than men to report that a health care provider didn’t understand their concerns, had assumed something without asking, talked down to them, made them feel uncomfortable, and didn’t believe them. When asked about how a doctor’s gender influences the likelihood of having a positive interaction, women were much more likely than men to report an own-gender doctor would be more likely to treat them with respect, understand their concerns, believe them, provide needed testing and treatments, make them feel comfortable, and ask appropriate questions instead of making assumptions. The asymmetry of responses between

³¹Figure 3 Panel A presents estimates based on the initially assigned choice set. Each point on this figure represents the fraction of respondents selecting an own-gender doctor among those who were assigned the relevant choice set as their first question. Figure 3 Panel B presents estimates which reflect the inferred willingness-to-pay of respondents for an own-gender doctor based on the full set of hypothetical choice questions asked of respondents.

³²The answer options for this question were: Not at all important, slightly important, moderately important, very important, and extremely important.

men and women supports the plausibility of the main finding: that doctor gender is an important factor in the evaluation of female patients more so than in the evaluation of male patients. Further, the survey evidence suggests differences in doctor behavior are likely an important driver of the main finding and highlights several possible ways male doctors—relative to their female counterparts—may be discounting the concerns and symptoms of female patients.

The survey evidence also helps us interpret the homophily evidence documented in the prior section. The survey evidence suggests that women are more likely to express a preference to see own-gender providers and more often select own-gender providers when given the choice. Additionally, the survey reveals gender differences in the intensity of preferences for own-gender providers, with women being willing to pay more to see an own-gender provider and more often reporting that doctor’s gender is an important factor when selecting a provider. These gender differences in preferences to see own-gender providers may reflect gender differences in patient expectations of having positive interactions with own-gender versus opposite-gender providers.

When interpreting the survey evidence, an important caveat to bear in mind is that the survey sample is not the same as the population reflected in the workers’ compensation insurance claims data.³³ While the survey sample is designed to be representative of the broader population of non-elderly workers, the workers’ compensation claims data cover only workers injured at work. Nevertheless, the survey provides suggestive evidence about potential mechanisms behind the main findings regarding the impact of gender concordance on medical evaluations and about the generalizability of—and causes behind—the observed gender homophily patterns in patient-provider matches.

8 Policy Counterfactuals and Discussion

We turn to investigating the potential policy implications of our findings. Our findings indicate that being evaluated by a female doctor increases evaluated disability and cash benefits for female patients, while having little impact on outcomes for male patients. At the same time, we document substantial gender gaps in this setting, where women are less likely to be evaluated as disabled and receive cash benefits than their male counterparts with the same observable characteristics. Thus, policies that increase the likelihood that female doctors evaluate female patients may increase evaluated disability and benefits for women and shrink documented gender gaps conditional on observables in this setting. We consider the impacts of two types of policies that would increase the prevalence of female doctors evaluating female patients: (i) increasing the share of doctors who are female and (ii) increasing gender homophily in patient-doctor matches.

First, consider both types of policies within our empirical setting—disputed claims with independent medical evaluations in the Texas workers’ compensation system. Independent medical evaluations—and policies surrounding these evaluations—are central in workers’ compensation programs, as roughly a third of claims with income benefits involve an independent medical exam. Our estimates imply that increasing the share of independent medical exams performed by female doctors from 17% to 50% would cause a 0.86 percentage point increase in female patients evaluated as disabled, closing 39.8% of the overall gender gap conditional on observables among disputed claims. An alternative—and more direct—way to increase the share of female patients with disputed claims evaluated by female doctors would be to change the doctor assignment process. For instance, patients with disputed claims could be randomly assigned to a doctor

³³The workers’ compensation administrative claims data do not include identifying information for claimants, so we could not have surveyed the exact population represented in the claims data.

of the same gender, rather than the current gender-blind random assignment process in which only 17% of female claimants are assigned female designated doctors. Such a policy would increase the share of female patients who receive cash benefits by 2.2 percentage points, closing the observed gender gap in benefit receipt.

More generally, we may be interested in the broader policy implications for gender gaps in workers' compensation insurance, beyond the subset of disputed claims with randomly assigned evaluators. We conduct back-of-the-envelope counterfactual analysis evaluating the potential impact of two types of policies on the gender gap: increasing the share of female doctors and increasing gender homophily between patients and doctors. There are several ways policy may influence the share of treating doctors who are female. For instance, some policies may encourage women to enter relevant medical specialties (e.g., residency hours caps (Wasserman 2019), while other policies may increase the share of female doctors accepting workers' compensation insurance patients (e.g., through outreach efforts or payment policy).³⁴ One can also imagine policies that increase gender homophily. For instance, gender homophily may increase if there were a reduction in barriers to sorting (e.g., providing patients a directory of workers' compensation treating doctors) or direct provision of information to claimants about the expected gains from sorting to same-gender doctors (e.g., informing female claimants of the results of this study). Our counterfactual analysis focuses on the reduced form impact of the gender composition of doctors and gender homophily on gender gaps. We note that an important area for future work is to understand how specific policy proposals impact the gender composition of doctors and patient sorting to doctors.

Below, we outline the setup for these back-of-the-envelope counterfactuals. In the discussion below, references to the "gender gap" pertain to the gender gap conditional on all available observable characteristics. To conduct this analysis, we combine our estimates on the effect of gender match from the randomized evaluations with our broader evidence on the gender gap and gender homophily in patient selection of treating doctors within workers' compensation insurance more generally. We note that this broader back-of-the-envelope counterfactual analysis involves extrapolation beyond the randomized evaluations, and thus the quantitative findings from this analysis should be interpreted with the appropriate caution.

Setup. We define some notation to describe this analysis. Let r_j be the share of patients of gender j who choose a female treating doctor, and let s denote the share of patients who are female. In this notation, the share of overall treating doctors who are female can be expressed as: $r \equiv r_m(1 - s) + r_f(s)$. We define relative gender homophily as the ratio of the share of female patients selecting female doctors to the share of male patients selecting female doctors, $Z \equiv \frac{r_f}{r_m}$. Let Gap_j denote the expected gender gap in benefit receipt—the percent reduction in the rate of benefit receipt for females relative to the analogous rate for males—when female claimants are evaluated by doctors of gender j . Using this notation, the overall gender gap in benefit receipt can be expressed as: $Gap \equiv Gap_m(1 - r_f) + Gap_f(r_f)$. We define the gender-match effect as the ratio of the gender gap when patients are evaluated by female doctors relative to the gender gap when patients are evaluated by male doctors: $X \equiv \frac{Gap_f}{Gap_m}$.

It is straightforward to show that we can express the gender gap as a function of the overall share of treating doctors who are female (r) and the degree of homophily in the market (Z):

$$Gap(r, Z) = Gap_m \times \left(1 - \frac{r}{\frac{1-s}{Z} + s}\right) + Gap_f \times \left(\frac{r}{\frac{1-s}{Z} + s}\right), \quad (7)$$

³⁴Refer to Goldin (2014) and Goldin and Katz (2016) for discussions of factors that make professions more female-welcoming.

where $Gap_f \equiv \frac{XGap^0}{r_f^0(X-1)+1}$, $Gap_m \equiv \frac{Gap^0}{r_f^0(X-1)+1}$, and Gap^0 and r_f^0 denote the overall gender gap and share of female patients selecting female doctors in the status quo, respectively. Appendix Section A presents this derivation in more detail. In the calculations that follow, we take the share of female patients (s) as given at the observed share in the workers' compensation insurance population: 38%.

We make two simplifying assumptions. First, we assume that patient-doctor gender match causes the same percent reduction in gender gaps in the likelihood of cash benefit receipt in broader workers' compensation insurance as it does within the set of claims with randomized evaluations. In other words, we assume the gender match effect (X) we estimate among the set of randomized evaluations applies more broadly within workers' compensation insurance. Second, we assume relative gender homophily in patient selection of treating doctors (Z) is constant with respect to the overall share of female treating doctors. For instance, female patients may select female doctors at Z times the rate that male patients select female doctors, but Z is constant as the share of doctors who are female varies.³⁵

Results. Using the relationship defined in Equation (7), Figure 4 illustrates the effect of varying one dimension holding the other fixed at the observed values in the status quo. In both panels, the vertical axis displays the expected gender gap in benefit receipt—the percent reduction in the rate at which females receive cash benefits relative to analogous rate for males with the same observable characteristics, while the horizontal axis displays either the share of female treating doctors (Panel A) or the relative gender homophily in patient-doctor matches (Panel B). For reference, each figure displays an “x” representing the observed combination in the status quo: a gender gap is 16.5% (i.e., females are 16.5% less likely to receive cash benefits than males with the same observables), the share of treating doctors who are female is 28.4%, and the degree of relative gender homophily is 1.058 (i.e., female patients select female doctors at 1.058 times the rate that male patients select female doctors).

Holding fixed the observed degree of gender homophily, Figure 4 Panel A plots the relationship between the gender gap and the share of female treating doctors. This figure indicates that increasing the share of female treating doctors by 21.6 percentage points—moving from 28.4% to parity—would lead to a 31.6% decrease in the observed gender gap. The observed degree of gender homophily in patient-doctor matches works to reinforce the effects of increasing gender diversity among doctors, with a 21.6 percentage point increase in the share of female treating doctors translating to a 22.4 percentage point increase in the share of female claimants who see female doctors. Additionally, Figure 4 Panel A can be used to determine the increase in the share of female doctors necessary to offset a given amount of the gender gap in cash benefit receipt. To offset a quarter of the gender gap in benefit receipt observed in the status quo, it would take a 16.9 percentage point increase in the share of treating doctors who are female.

Figure 4 Panel B plots the relationship between the gender gap and the degree of gender homophily, fixing the gender composition of doctors as observed in the status quo. An increase in relative gender homophily from the observed level to two (i.e., where female patients choose female doctors at twice the rate that male patients choose female doctors), would lead to a 16.6% decrease in the observed gender gap. Holding all else equal, female patients would need to choose female doctors 2.7 times as often as male patients choose female doctors to offset a quarter of the observed gender gap.

More broadly, we can characterize the trade-offs between policies that may increase gender diversity

³⁵While it would be straightforward to extend these calculations to allow gender homophily to vary with the share of doctors who are female, we avoid doing this for a few reasons. First, whether Z and r are correlated—and the direction of any correlation—is ex ante unclear. Second, we cannot credibly estimate how Z varies with r . While Section 6 presents estimates of relative gender homophily across the workers' compensation insurance system in Texas, there is no plausibly exogenous variation in the availability of female doctors across geography to identify how relative gender homophily varies with the share of doctors who are female.

among doctors and policies that increase sorting of patients to same-gender doctors. Figure 5 illustrates level curves of Equation (7), which characterize the combinations of gender composition in the doctor workforce and gender homophily patient-doctor matches that would result in a given value of the gender gap. The combination of conditions in the status quo is represented as “x” in this figure. There are at least two important qualitative properties worth noting. First, the level curves are negatively sloped. This indicates that a given reduction in the gender gap can be achieved by trading-off increases in the share of doctors who are female and increases in relative gender homophily. Second, the level curves are convex relative to the origin. This convexity reflects the fact that the inputs—the share of doctors who are female and relative gender homophily—are complementary in reducing the gender gap.

Suppose we are interested in understanding the change in conditions needed to reduce the gender gap from 16.5% to 15%. This reduction in the gender gap could be accomplished by increasing the share of doctors female to 34.5%, holding homophily fixed at 1.058. Alternatively, it could be accomplished by increasing homophily to 1.50, holding the share of female doctors fixed at 28.4%. More generally, the combinations of conditions that would lead to the same closure of the gap are depicted by the level curve corresponding to 15%, where the difference between this curve and the observed conditions “x” indicates the changes in conditions necessary to achieve this closure. Reducing the gender gap to 15% could be accomplished by a convex combinations of changes in the gender composition of treating doctors and relative gender homophily—for example, by increasing gender homophily to 1.25 and the share of treating doctors who are female to 31.3%, or increasing gender homophily to 1.1 and the share of treating doctors who are female to 33.7%.

9 Conclusion

This paper investigates how male and female doctors differ in their medical evaluations of male and female patients. Leveraging information on injured workers randomly assigned to doctors for medical evaluations in the Texas workers’ compensation insurance system, we show that being evaluated by a female doctor rather than a male doctor increases the likelihood that female claimants are evaluated as disabled and increases subsequent cash disability benefits females receive. In contrast, we show the gender of the evaluating doctor does not impact evaluated disability or subsequent cash benefits for male claimants. These patterns hold for various subgroups and do not seem to be explained by other characteristics of the claimants or the evaluating doctors. Our findings indicate that gender of evaluating doctors is an important determinant of gender gaps in this setting, where the estimates imply that having female doctors evaluate patients—rather than male doctors—eliminates the observed gender gap in the likelihood of being evaluated as disabled conditional on observables.

To place these estimates within the broader context, we use data on all workers’ compensation claims to assess male-female benefit differences more broadly and to examine patient-doctor sorting when patients have the ability to select their treating doctors. We find that female claimants are 16.5% less likely to receive cash disability benefits than male claimants are after we control for a rich set of baseline claimant and injury characteristics. Relative to comparable male claimants, female claimants are about 6% more likely to choose a female treating doctor when they are allowed to choose their own doctor, which is consistent with female patients having a preference for female doctors on average relative to male patients’ preference for female doctors. Results from a supplemental survey we conduct suggest that women—relative to men—more often have negative health care experiences, expect better treatment from own-gender doctors, and have stronger preferences to see own-gender doctors.

One implication of our findings is that increasing the share of doctors who are female may increase assessed disability and cash benefits for women, while having little impact on outcomes for men. Extrapolating from our estimates, we find that increasing the share of independent medical evaluations performed by female doctors from 17% to 50% would increase the share of female patients evaluated as disabled, closing 39.8% of the overall gender gap conditional on observables among claimants with independent medical evaluations. To analyze the broader implications of our findings, we conduct back-of-the-envelope counterfactuals combining the broader evidence on gender disparities in workers' compensation insurance with the estimated gender-match effects among claimants with randomized independent medical evaluations. This analysis characterizes the combination of policies—aimed at either increasing the share of doctors who are female or increasing gender homophily in patient-doctor matches—that would result in a given reduction in the gender gap. These calculations reveal that these types of policies can substantially shrink gender gaps in evaluated disability and work to complement one another in doing so.

More broadly, our findings highlight the importance of gatekeeper discretion in public programs and inequities that can arise from this discretion. Our findings illustrate that gender match between program gatekeepers and claimants can have important implications for the distribution of program benefits. While our counterfactual analysis focuses on policies that shift the composition of evaluators, policymakers may also want to consider broader policies aimed at reducing the impact of discretion on evaluations—such as, requiring multiple assessments for high-stakes evaluations, mandating evaluators receive de-biasing training, or encouraging the use of claimant advocates. Analyzing the impacts of such policies is an important area for future work.

Our results also inform broader literatures that study the differential treatment of male and female patients within the health care system and the impact of the gender of authority figures and evaluators on females more broadly. Despite a growing body of evidence documenting that male and female patients receive different treatment in the health care system for reasons that appear unrelated to underlying health needs, determining if structural factors of the health care system play a role in these differences is difficult because the assignment of patients to doctors is typically non-random. Our study suggests that differential evaluations of female patients by male and female doctors likely play a role in differential outcomes for male and female patients. A natural corollary of these implications is that the under-representation of female doctors is likely an important contributing factor behind observed gender disparities in health care settings. More broadly, our findings highlight the potential importance of gender match in settings where authority figures evaluate women.

References

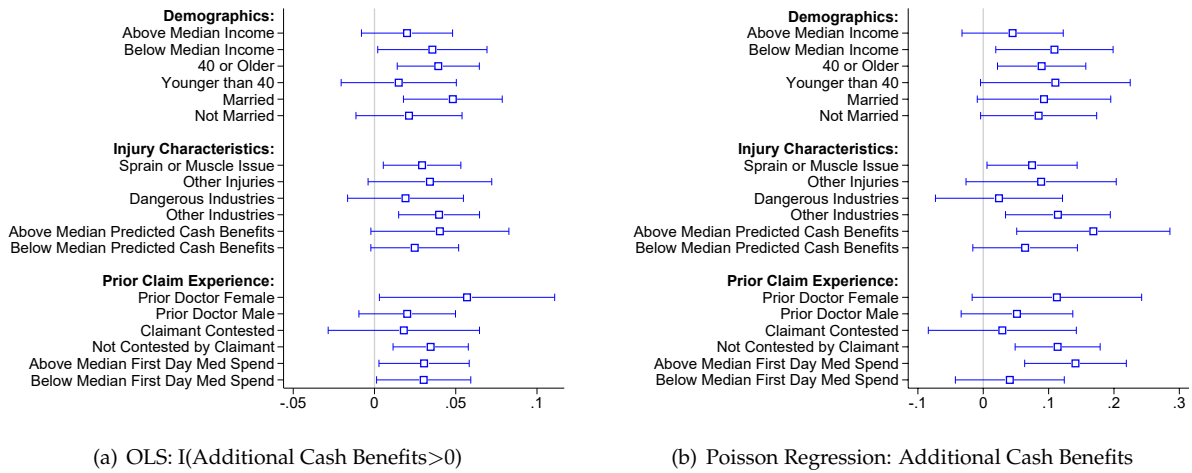
- Abrevaya, Jason and Daniel S. Hamermesh.** 2012. "Charity and Favoritism in the Field: Are Female Economists Nicer (To Each Other)?" *The Review of Economics and Statistics*, 94(1): 202–207.
- Alsan, Marcella, Owen Garrick, and Grant Graziani.** 2019. "Does Diversity Matter for Health? Experimental Evidence from Oakland." *American Economic Review*, 109(12): 4071–4111.
- Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik.** 2015. "The Effect of Teacher Gender on Student Achievement in Primary School." *Journal of Labor Economics*, 33(1): 63–89.
- Bertrand, Marianne, Sandra E Black, Sissel Jensen, and Adriana Lleras-Muney.** 2018. "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway." *The Review of Economic Studies*, 86(1): 191–239.

- Biasi, Barbara and Heather Sarsons.** 2020. "Flexible Wages, Bargaining, and the Gender Gap." National Bureau of Economic Research Working Paper 27894.
- Cabral, Marika and Marcus Dillender.** 2020. "Disparities in Health Care and Medical Evaluations by Gender: A Review of Evidence and Mechanisms." *AEA Papers and Proceedings*.
- Card, David, Stefano Della Vigna, Patricia Funk, and Nagore Iriberry.** 2019. "Are Referees and Editors in Economics Gender Neutral?*" *The Quarterly Journal of Economics*, 135(1): 269–327.
- Carrell, Scott E., Marianne E. Page, and James E. West.** 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *The Quarterly Journal of Economics*, 125(3): 1101–1144.
- Casarico, Alessandra and Salvatore Lattanzio.** 2019. "What Firms Do: Gender Inequality in Linked Employer-Employee Data." University of Cambridge Working Paper.
- Centers for Medicare & Medicaid Services (CMS).** 2019. "National Plan and Provider Enumeration System." Centers for Medicare & Medicaid Services. <https://npiregistry.cms.hhs.gov/>. (Date accessed: 07/2019).
- Centers for Medicare & Medicaid Services (CMS).** 2021. "Doctors and Clinicians National Downloadable File: Performance Year 2019." Centers for Medicare & Medicaid Services. <https://data.cms.gov/provider-data/dataset/mj5m-pzi6>. (Date accessed: 08/2021).
- Chandra, Amitabh and Douglas O Staiger.** 2010. "Identifying Provider Prejudice in Healthcare." National Bureau of Economic Research Working Paper 16382.
- Chen, Esther H, Frances S Shofer, Anthony J Dean, Judd E Hollander, William G Baxt, Jennifer L Robey, Keara L Sease, and Angela M Mills.** 2008. "Gender disparity in analgesic treatment of emergency department patients with acute abdominal pain." *Academic Emergency Medicine*, 15(5): 414–418.
- Cullen, Zoë B and Ricardo Perez-Truglia.** 2019. "The Old Boys' Club: Schmoozing and the Gender Gap." National Bureau of Economic Research Working Paper 26530.
- Côté, Daniel and Marie-France Coutu.** 2010. "A critical review of gender issues in understanding prolonged disability related to musculoskeletal pain: how are they relevant to rehabilitation?" *Disability and Rehabilitation*, 32(2): 87–102. PMID: 21495273.
- Daly, Mary C, Brian Lucking, Jonathan A Schwabish, et al.** 2013. "The Future of Social Security Disability Insurance." *FRBSF Economic Letter*, 17.
- Department of Veterans Affairs.** 2010. "Review of Combat Stress in Women Veterans Receiving VA Health Care and Disability Benefits." Office of Inspector General.
- Dobbie, Will and Jae Song.** 2015. "Debt Relief and Debtor Outcomes: Measuring the Effects of Consumer Bankruptcy Protection." *American Economic Review*, 105(3): 1272–1311.
- Dupas, Pascaline and Radhika Jain.** 2021. "Women Left Behind: Gender Disparities in Utilization of Government Health Insurance in India." National Bureau of Economic Research Working Paper 28972.
- Goldin, Claudia.** 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review*, 104(4): 1091–1119.
- Goldin, Claudia and Lawrence F Katz.** 2016. "A Most Egalitarian Profession: Pharmacy and the Evolution of a Family-Friendly Occupation." *Journal of Labor Economics*, 34(3): 705–746.
- Greenwood, Brad N., Seth Carnahan, and Laura Huang.** 2018. "Patient–physician gender concordance and increased mortality among female heart attack patients." *Proceedings of the National Academy of Sciences - PNAS*, 115(34): 8569–8574.

- Hernandez, Adrian F., Gregg C. Fonarow, Li Liang, Sana M. Al-Khatib, Lesley H. Curtis, Kenneth A. LaBresh, Clyde W. Yancy, Nancy M. Albert, and Eric D. Peterson.** 2007. "Sex and Racial Differences in the Use of Implantable Cardioverter-Defibrillators Among Patients Hospitalized With Heart Failure." *JAMA*, 298(13): 1525–1532.
- Hill, Andrew J., Daniel B. Jones, and Lindsey Woodworth.** 2020. "A Doctor Like Me: Physician-Patient Race-Match and Patient Outcomes." Working Paper.
- Hoffmann, Diane E. and Anita J. Tarzian.** 2001. "The Girl Who Cried Pain: A Bias against Women in the Treatment of Pain." *The Journal of Law, Medicine & Ethics*, 28(4_suppl): 13–27.
- Jagsi, Reshma, Kent A Griffith, Rochelle A DeCastro, and Peter Ubel.** 2014. "Sex, role models, and specialty choices among graduates of US medical schools in 2006–2008." *Journal of the American College of Surgeons*, 218(3): 345–352.
- Kaiser Family Foundation.** 2019. "Professionally Active Physicians by Gender."
- Lim, Jaegeum and Jonathan Meer.** 2017. "The Impact of Teacher–Student Gender Matches: Random Assignment Evidence from South Korea." *The Journal of Human Resources*, 52(4): 979–997.
- Lim, Jaegeum and Jonathan Meer.** 2019. "Persistent Effects of Teacher–Student Gender Matches." *The Journal of Human Resources*, 55(3): 809–835.
- Low, Hamish and Luigi Pistaferri.** 2019. "Disability Insurance: Error Rates and Gender Differences." National Bureau of Economic Research Working Paper 26513.
- Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand.** 2013. "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." *American Economic Review*, 103(5): 1797–1829.
- Maida, Agata and Andrea Weber.** 2019. "Female Leadership and Gender Gap within Firms: Evidence from an Italian Board Reform." *ILR Review*, 0(0): 0019793920961995.
- McDevitt, Ryan C. and James W. Roberts.** 2014. "Market structure and gender disparity in health care: preferences, competition, and quality of care." *The RAND Journal of Economics*, 45(1): 116–139.
- Muralidharan, Karthik and Ketki Sheth.** 2016. "Bridging Education Gender Gaps in Developing Countries: The Role of Female Teachers." *The Journal of Human Resources*, 51(2): 269–297.
- Murdoch, Maureen, Michele Roxanne Spoot, Nina Aileen Sayer, Shannon Marie Kehle-Forbes, and Siamak Noorbaloochi.** 2021. "Reversals in Initially Denied Department of Veterans Affairs' PTSD Disability Claims after 17 Years: A Cohort Study of Gender Differences." *BMC Women's Health*, 21(70): 1–8.
- National Conference of State Legislatures.** 2010. "2010 Legislation on Workers' Compensation."
- Pelletier, Roxanne, Karin H Humphries, Avi Shimony, Simon L Bacon, Kim L Lavoie, Doreen Rabi, Igor Karp, Meytal Avgil Tsadok, and Louise Pilote.** 2014. "Sex-related differences in access to care among patients with premature acute coronary syndrome." *Canadian Medical Association journal (CMAJ)*, 186(7): 497–504.
- Pietropaoli, Anthony P., Laurent G. Glance, David Oakes, and Susan G. Fisher.** 2010. "Gender differences in mortality in patients with severe sepsis or septic shock." *Gender Medicine*, 7(5): 422–437.
- Samulowitz, Anke, Ida Gremyr, Erik Eriksson, and Gunnel Hensing.** 2018. "'Brave Men' and 'Emotional Women': A Theory-Guided Literature Review on Gender Bias in Health Care and Gendered Norms towards Patients with Chronic Pain." *Pain research & management*, 2018: 1–14.
- Sarsons, Heather.** 2019. "Interpreting Signals in the Labor Market: Evidence from Medical Referrals." *Working Paper*.

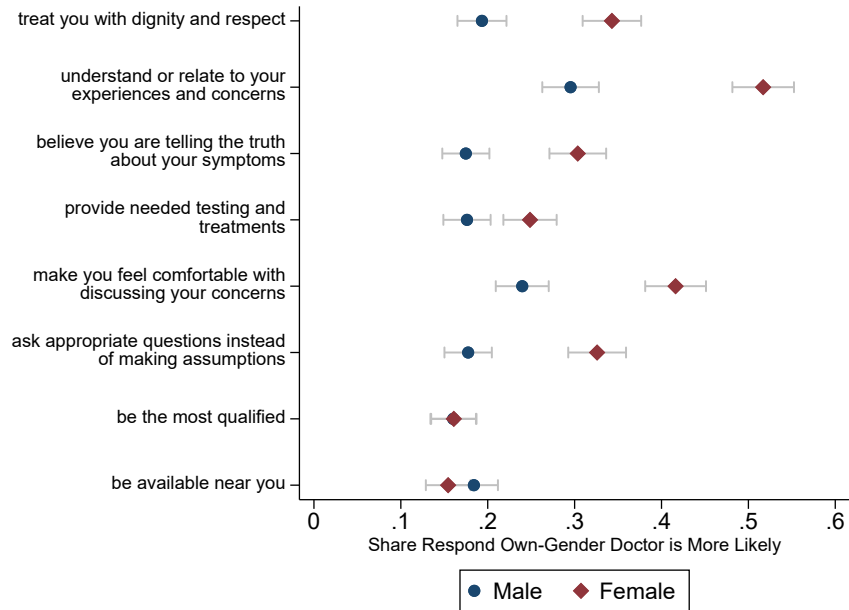
- Schnell, Molly and Janet Currie.** 2018. "Addressing the Opioid Epidemic: Is There a Role for Physician Education?" *American Journal of Health Economics*, 4(3): 383–410.
- Schulman, Kevin A, Jesse A Berlin, William Harless, Jon F Kerner, Shyrl Sistrunk, Bernard J Gersh, Ross Dube, Christopher K Taleghani, Jennifer E Burke, Sankey Williams, et al.** 1999. "The effect of race and sex on physicians' recommendations for cardiac catheterization." *New England Journal of Medicine*, 340(8): 618–626.
- Texas Board of Chiropractic Examiners.** 2020. "Texas Board of Chiropractic Examiners Doctors of Chiropractic License Data (2013-2017)." Texas Board of Chiropractic Examiners.
- Texas Department of Insurance (TDI).** 2018a. "Texas Workers' Compensation Insurance Designated Doctor Exam Data (2012-2018)." Texas Department of Insurance.
- Texas Department of Insurance (TDI).** 2018b. "Texas Workers' Compensation Insurance Linked Medical and Cash Benefit Data (2013-2018)." Texas Department of Insurance.
- Texas Department of Insurance (TDI).** 2020. "2020 Workers' Compensation Network Report Card Results."
- Trabanino, Shawn.** 2020. "Health, Law, And Ethnicity: The Disability Administrative Law Judge and Health Disparities for Disadvantaged Populations." *Calif. L. Rev.*, 108: 2079.
- Tunks, Eldon Ed, Anthony Ed Bellissimo, and Ranjan Ed Roy.** 1990. *Chronic Pain: Psychosocial Factors in Rehabilitation*. Robert E Krieger Publishing Co.
- Vaccarino, Viola, Saif S. Rathore, Nanette K. Wenger, Paul D. Frederick, Jerome L. Abramson, Hal V. Barron, Ajay Manhapra, Susmita Mallik, and Harlan M. Krumholz.** 2005. "Sex and Racial Differences in the Management of Acute Myocardial Infarction, 1994 through 2002." *New England Journal of Medicine*, 353(7): 671–682. PMID: 16107620.
- Wasserman, Melanie.** 2019. "Hours Constraints, Occupational Choice, and Gender: Evidence from Medical Residents." Working Paper.
- Weisse, Carol S, Paul C Sorum, and Rachel E Dominguez.** 2003. "The influence of gender and race on physicians' pain management decisions." *The Journal of Pain*, 4(9): 505–510.
- Zeltzer, Dan.** 2020. "Gender Homophily in Referral Networks: Consequences for the Medicare Physician Earnings Gap." *American Economic Journal: Applied Economics*, 12(2): 169–97.

Figure 1: Heterogeneity: Estimate on Interaction of Female Doctor and Female Claimant by Subgroup



Notes: Each marker is the coefficient on the interaction of the female doctor and female claimant indicator variables from separate regressions of Equation (1) for the specified sample of claimants. All regressions control for a female doctor indicator variable, a female claimant indicator variable, credential-by-county fixed effects, exam year fixed effects, and injury year fixed effects. The dependent variables are as indicated in the figure: an indicator for receiving any additional cash benefits (Panel A) and (normalized) additional cash benefits received (Panel B). As described in Section 3, the presence of wage and industry information is related to the receipt of cash benefits, and thus these variables are only available for a subset of claimants. The 95-percent confidence intervals displayed along with the coefficient estimates are calculated using standard errors clustered at the doctor level. The sample includes claims occurring from 2013 to 2017 that had an independent medical exam by the end of 2017 and that have non-missing values for the specified characteristic.

Figure 2: Survey: “Thinking about your experiences getting health care for yourself, which doctor—male or female—would be more likely to...” [answer options male, female, equally likely]



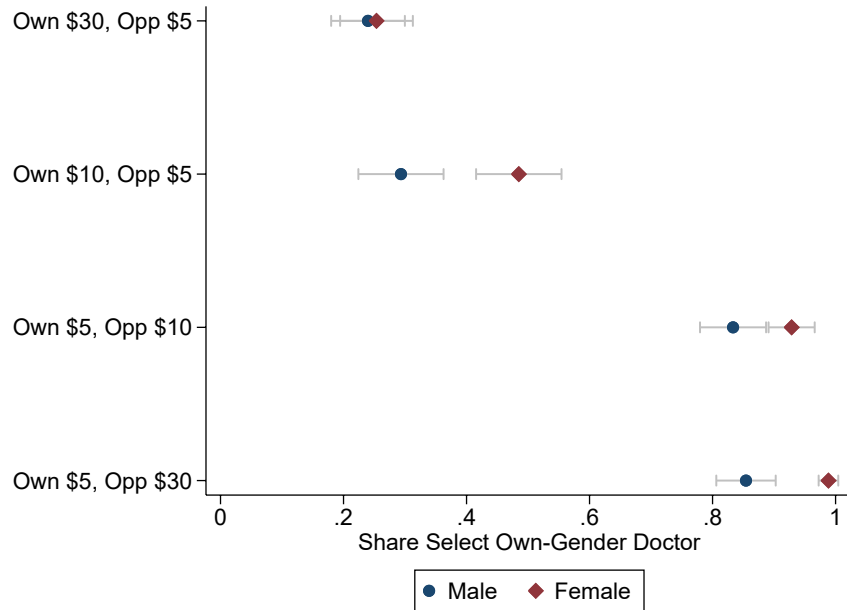
(a) Share that Select Own-Gender Doctor



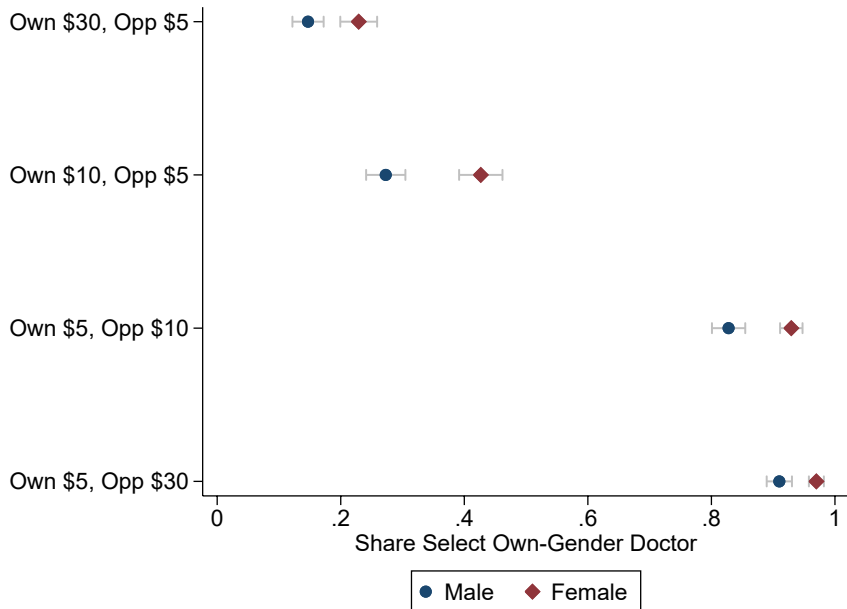
(b) Share that Do Not Select Opposite-Gender Doctor

Notes: The figure above shows the means and associated 95% confidence intervals for survey responses to the indicated question by respondent gender. The survey was conducted by Qualtrics and included 1,519 adults ages 30 to 64. Eligibility was restricted to individuals who reported working at some point in the last 12 months and individuals who self-identified as either male (755 respondents) or female (764 respondents). See Section 7 for more detail on the survey. Table 11 reports raw and regression-adjusted differences in means for these survey questions and the associated p-values.

Figure 3: Survey: Share Selecting an Own-Gender Doctor in Hypothetical Choice Questions When Varying Co-Pay Differential



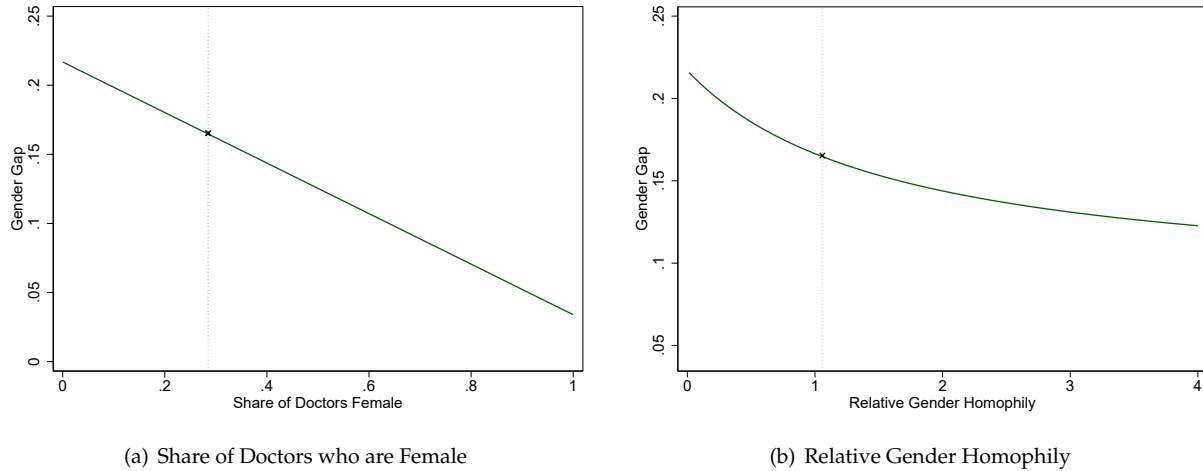
(a) Based on Initial Hypothetical Choice Question



(b) Based on Full Set of Hypothetical Choice Questions

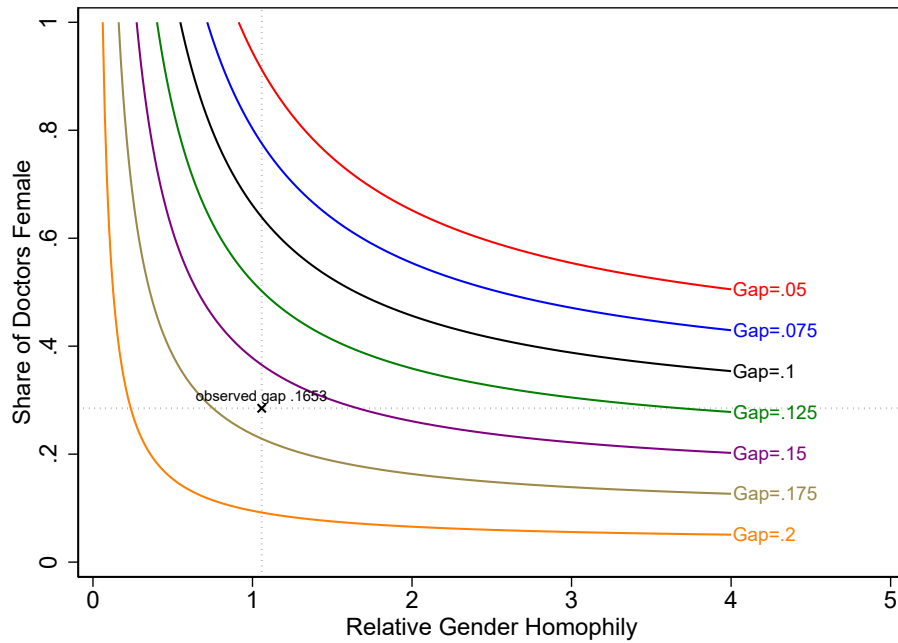
Notes: The figure above shows the share selecting an own-gender doctor for each co-pay differential in hypothetical choice questions and the 95% confidence intervals by respondent gender. The survey was conducted by Qualtrics and included 1,519 adults ages 30 to 64. Eligibility was restricted to individuals who reported working at some point in the last 12 months and individuals who self-identified as either male (755 respondents) or female (764 respondents). See Section 7 for more detail on the survey. Table 12 reports raw and regression-adjusted differences in means for these survey questions and the associated p-values.

Figure 4: Counterfactual Policy Analysis: Partial Effects of Varying Share Doctors Female or Gender Homophily on Gender Gaps



Notes: This figure displays the results from policy counterfactual analysis described in Section 8. This analysis combines the broader findings regarding the gender gap and relative gender homophily in workers' compensation more generally, with the estimated effects of gender match among claims with randomized evaluations. This analysis draws on Equation (7), which relates the gender gap (conditional on observable characteristics) to the share of doctors who are female and relative gender homophily in patient-doctor matches. For the purposes of this figure, the gender gap represents the percent reduction in the likelihood of cash benefits for female patients relative to male patients with the same observable characteristics. The point indicating the observed values in the status quo is indicated with an "x" in each panel, where the gender gap is 16.5% (i.e., females are 16.5% less likely to receive benefits than males with the same observables), the share of treating doctors who are female is 0.285, and the degree of relative gender homophily is 1.058 (i.e., female patients select female doctors at 1.058 times the rate that male patients select female doctors). Panel A displays the partial effects of varying the share of doctors who are female, holding relative gender homophily fixed. Panel B displays the partial effects of varying the degree of relative gender homophily, holding the share of doctors who are female fixed.

Figure 5: Counterfactual Policy Analysis: Effect of Share Doctors Female and Gender Homophily on Gender Gaps



Notes: This figure displays the results from policy counterfactual analysis described in Section 8. This analysis combines the broader findings regarding the gender gap and relative gender homophily in workers' compensation more generally, with the estimated effects of gender match among claims with randomized evaluations. This figure shows the level curves of Equation (7), which relates the gender gap (conditional on observables) to the share of doctors who are female and relative gender homophily in patient-doctor matches. For the purposes of this figure, the gender gap represents the percent reduction in the likelihood of cash benefits for female patients relative to male patients with the same observable characteristics. The point indicating the observed values in the status quo is indicated with an "x", where the gender gap is 16.5% (i.e., females are 16.5% less likely to receive benefits than males with the same observables), the share of treating doctors who are female is 0.285, and the degree of relative gender homophily is 1.058 (i.e., female patients select female doctors at 1.058 times the rate that male patients select female doctors).

Table 1: Summary Statistics

	All Claimants		Baseline Sample Claimants with Randomized Evaluations	
	Male (1)	Female (2)	Male (3)	Female (4)
Age	40.0	43.0	44.4	47.0
ED Claim	0.284	0.230	0.368	0.306
First-Day Medical Spending	639	442	1,305	807
Three-Month Medical Spending	2,013	1,497	6,273	4,983
Receives Income-Replacement Benefits within 1 Year of Injury	0.215	0.162	0.821	0.783
Receives Impairment Benefits within 1 Year of Injury	0.076	0.056	0.473	0.445
Injury Type:				
Contusion	0.090	0.150	0.051	0.082
Fracture	0.052	0.037	0.126	0.097
Laceration	0.145	0.090	0.052	0.018
Muscle Issue	0.193	0.221	0.308	0.337
Sprain	0.224	0.260	0.267	0.303
Pre-Injury Weekly Wage	911	687	957	719
Industry:				
Agriculture/Forestry/Fishing/Hunting	0.015	0.006	0.015	0.007
Arts/Entertainment/Accommodation/Food Services	0.043	0.079	0.027	0.073
Finance/Real Estate/Professional Services	0.062	0.073	0.074	0.094
Health Care/Educational Services	0.081	0.413	0.057	0.329
Manufacturing	0.142	0.066	0.171	0.099
Mining/Utilities/Construction	0.200	0.018	0.239	0.027
Public Administration/Other Services	0.262	0.197	0.213	0.196
Wholesale Trade/Retail Trade/Transportation	0.195	0.149	0.204	0.176
Has Female Treating Doctor	0.275	0.299	0.220	0.239
Has Female Designated Doctor	-	-	0.173	0.177
Receives Additional Cash Benefits within 1 Year of Exam	-	-	0.628	0.601
Amount of Additional Cash Benefits within 1 Year of Exam	-	-	6,567	5,729
Percent of Sample	62%	38%	66%	34%

Notes: This table displays means for all 1,076,759 claims occurring from 2013 to 2017 and for the 71,366 claims occurring from 2013 to 2017 that had an independent medical exam by the end of 2017. As described in Section 3, the presence of wage and industry information is related to the receipt of cash benefits, and thus these variables are only available for a subset of claimants.

Table 2: Doctor Characteristics

	Designated Doctors (1)	Texas Doctors Treating Injured Workers (2)	All Texas Doctors (3)	All U.S. Doctors (4)
MDs/DOs	60%	85%	91%	92%
Gender:				
Female	19%	22%	34%	34%
Male	81%	78%	66%	66%
Specialty:				
Internal or Family Medicine	38%	47%	37%	38%
Orthopedics	16%	21%	3%	3%
Other	46%	31%	60%	59%
Top 25 Medical School	16%	17%	19%	19%
DCs	40%	15%	9%	8%
Gender:				
Female	18%	18%	28%	28%
Male	82%	82%	72%	72%
N	1,461	5,587	64,700	1,022,197

Notes: Column 1 contains information on doctors who performed designated doctor exams from 2013 to 2017 in doctor exam data from the Texas Department of Insurance. Column 2 contains information on doctors who treated injured workers from 2013 to 2017 in the Texas workers' compensation insurance medical data. Columns 3 and 4 contain information on all doctors in the National Plan and Provider Enumeration System data through 2018. Doctors are identified using a unique identifier from the Texas Department of Insurance in column 1 and using NPI in other columns. Information on age and specialty comes from the National Plan and Provider Enumeration System data. Information on medical schools attended comes from the Medicare Physician Compare File. Information on medical school rankings comes from schools' average research ranking in the U.S. News and World Report from 2010 to 2017 as reported in Schnell and Currie (2018). If an observation is missing information for a characteristic, that observation is excluded from the percent calculation so total percents sum to 100.

Table 3: Balance

	Female Doctor X Female Claimant			Female Doctor			Female Claimant		
	Coeff	Std Err	p-value	Coeff	Std Err	p-value	Coeff	Std Err	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age	-0.133	(0.272)	[0.625]	0.099	(0.148)	[0.502]	2.663	(0.108)	[<0.001]
ED Claim	-0.007	(0.010)	[0.495]	0.005	(0.007)	[0.461]	-0.038	(0.004)	[<0.001]
Log(First-Day Medical Spending)	0.036	(0.030)	[0.236]	-0.028	(0.020)	[0.167]	-0.199	(0.013)	[<0.001]
Log(Med Spending Prior to Exam)	-0.001	(0.021)	[0.953]	-0.002	(0.014)	[0.909]	-0.163	(0.010)	[<0.001]
Receives Income-Replacement Benefits Prior to Exam	0.011	(0.008)	[0.187]	0.002	(0.005)	[0.727]	-0.040	(0.004)	[<0.001]
Log(Weeks from Injury to Exam)	-0.012	(0.009)	[0.166]	0.005	(0.006)	[0.439]	0.001	(0.004)	[0.800]
Injury Type:									
Contusion	-0.001	(0.005)	[0.861]	0.004	(0.003)	[0.157]	0.030	(0.002)	[<0.001]
Fracture	0.005	(0.007)	[0.525]	-0.002	(0.004)	[0.607]	-0.022	(0.003)	[<0.001]
Laceration	-0.001	(0.004)	[0.794]	-0.000	(0.003)	[0.938]	-0.029	(0.002)	[<0.001]
Muscle Issue	0.009	(0.011)	[0.412]	-0.005	(0.006)	[0.387]	0.021	(0.004)	[<0.001]
Sprain	-0.012	(0.009)	[0.223]	0.003	(0.006)	[0.666]	0.020	(0.004)	[<0.001]

Notes: This table displays estimates of the coefficients on the female doctor indicator variable, the female claimant indicator variable, and the interaction of the female doctor and female claimant indicator variables from OLS regressions of Equation (1) that control for credential-by-county fixed effects, exam year fixed effects, and injury year fixed effects. Each row represents a separate regression with the dependent variable as indicated in the table. Columns 1, 4, and 7 display the coefficient estimates, columns 2, 5, and 8 display standard errors clustered at the doctor level, and columns 3, 6, and 9 display p-values. In each specification, the sample includes claims occurring from 2013 to 2017 that had an independent medical exam by the end of 2017 and that have non-missing values for the given dependent variable.

Table 4: Effect on Benefit Receipt

	OLS: I(Additional Cash Benefits > 0)			Poisson Regression: Additional Cash Benefits		
	(1)	(2)	(3)	(4)	(5)	(6)
Female Doctor X Female Claimant	0.030 (0.010) [0.003]	0.029 (0.010) [0.005]		0.085 (0.030) [0.005]	0.079 (0.031) [0.011]	
Female Doctor	-0.0002 (0.008) [0.984]		0.030 (0.011) [0.005]	0.031 (0.026) [0.221]		0.116 (0.034) [0.001]
Female Claimant	-0.031 (0.004) [<0.001]	-0.030 (0.004) [<0.001]		-0.139 (0.013) [<0.001]	-0.134 (0.013) [<0.001]	
Sample	All Claimants	All Claimants	Female Claimants	All Claimants	All Claimants	Female Claimants
Doctor Fixed Effects		x			x	
Mean of Dep. Var.	0.619	0.619	0.601	6,284	6,284	5,729
Mean of Dep. Var. for Females Evaluated by Male Doctors	0.596	0.596	0.596	5,627	5,627	5,627
N	71,366	71,366	24,054	71,366	71,366	24,054

Notes: This table displays estimates of the coefficients on the female doctor indicator variable, the female claimant indicator variable, and the interaction of the female doctor and female claimant indicator variables from regressions of Equations (1), (2), and (3) that control for credential-by-county fixed effects, exam year fixed effects, and injury year fixed effects. Each column represents a separate regression with the dependent variable as indicated in the table. The sample includes claims occurring from 2013 to 2017 that had an independent medical exam by the end of 2017. Standard errors clustered at the doctor level are reported in parentheses, and p-values are reported in brackets.

Table 5: Other Patient Characteristics

	Dependent Variable: I(Additional Cash Benefits > 0)								
	Demographics			Injury Characteristics			Prior Medical Experience		
	Income < Median	Age ≥ 40	Married	Sprain or Muscle Issue	Dangerous Industry	Predicted Cash Benefits > Median	Prior Doctor Female	Claimant Contested	First Day Med Spend > Median
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
X * Female Doctor	0.006 (0.010) [0.531]	-0.001 (0.010) [0.959]	0.011 (0.012) [0.364]	0.014 (0.012) [0.235]	-0.002 (0.009) [0.809]	-0.013 (0.010) [0.202]	0.006 (0.014) [0.676]	0.002 (0.012) [0.884]	-0.004 (0.010) [0.660]
Female Doctor	0.004 (0.009) [0.613]	0.010 (0.010) [0.276]	0.003 (0.010) [0.781]	0.001 (0.012) [0.934]	0.011 (0.009) [0.223]	0.017 (0.009) [0.065]	0.013 (0.009) [0.152]	0.010 (0.008) [0.213]	0.012 (0.009) [0.178]
X	-0.002 (0.004) [0.555]	0.057 (0.004) [<0.001]	-0.004 (0.005) [0.440]	0.046 (0.005) [<0.001]	0.045 (0.004) [<0.001]	-0.006 (0.004) [0.177]	0.006 (0.006) [0.347]	0.042 (0.006) [<0.001]	0.011 (0.004) [0.006]

Notes: This table displays estimates of the coefficients on the female doctor indicator variable, the indicated claim or claimant characteristic, and the interaction of the female doctor indicator and the indicated claim or claimant characteristic—estimates from regressions of Equation (1) replacing the female claimant indicator with the characteristic indicated in the column. These regressions additionally control for credential-by-county fixed effects, exam year fixed effects, and injury year fixed effects. Each column represents a separate regression with the dependent variable being I(Additional Cash Benefits > 0). As described in Section 3, the presence of wage and industry information is related to the receipt of cash benefits. The sample includes claims occurring from 2013 to 2017 that had an independent medical exam by the end of 2017 with non-missing information for the indicated characteristic. Standard errors clustered at the doctor level are reported in parentheses, and p-values are reported in brackets.

Table 6: Other Doctor Characteristics

	Dependent Variable: DoctorXFemale-Claimant Fixed Effects Estimated Based on All Exams						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Doctor	0.026 (0.010) [0.013]	0.026 (0.010) [0.013]	0.026 (0.010) [0.012]	0.025 (0.011) [0.018]	0.027 (0.011) [0.011]	0.025 (0.011) [0.018]	0.026 (0.011) [0.020]
Doctor Credential (MD/DO)		0.007 (0.009) [0.420]	0.006 (0.009) [0.516]	0.006 (0.013) [0.616]			0.0003 (0.0144) [0.981]
Doctor Degree from Top 25 Medical School			0.009 (0.025) [0.715]				0.009 (0.025) [0.729]
Doctor Specialty: Internal or Family Medicine				0.005 (0.017) [0.771]			0.008 (0.017) [0.650]
Doctor Specialty: Orthopedics				-0.007 (0.018) [0.713]			-0.004 (0.018) [0.820]
Doctor Experience					0.0003 (0.0003) [0.346]		0.0003 (0.0004) [0.466]
Doctor State of Birth Texas						0.002 (0.009) [0.826]	0.005 (0.009) [0.548]
N	917	917	917	904	878	887	878

Notes: This table displays estimates of the coefficients on the female doctor indicator variable and on other doctor characteristics from a regression of doctors' female-claimant coefficients on doctor characteristics. To obtain the female-claimant coefficients for doctors used as the dependent variable, we first regress the indicator variable for doctors assessing claimants as having continued disability on injury year fixed effects, exam year fixed effects, doctor fixed effects, and separate doctor-by-female-claimant interaction terms for each of the 917 doctors with at least ten exams in the sample. The doctor-by-female-claimant fixed effects estimated from this regression are used as the dependent variable in the analysis above. Each column represents a separate regression weighted by the number of exams doctors performed. The sample includes designated doctors who performed at least ten independent medical exams from 2013 to 2017. Standard errors are reported in parentheses, and p-values are reported in brackets.

Table 7: Effect on Indicator for Receiving Additional Benefits after Controlling for Additional Claimant and Injury Characteristics

	Dependent Variable: I(Additional Cash Benefits > 0)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Doctor X Female Claimant	0.030 (0.010) [0.003]	0.026 (0.010) [0.010]	0.026 (0.010) [0.010]	0.027 (0.010) [0.008]	0.026 (0.010) [0.013]	0.026 (0.010) [0.012]	0.026 (0.010) [0.013]
Female Doctor	-0.000 (0.008) [0.984]	0.001 (0.008) [0.887]	0.001 (0.008) [0.884]	0.001 (0.008) [0.943]	0.001 (0.008) [0.872]	0.001 (0.008) [0.889]	0.001 (0.008) [0.872]
Female Claimant	-0.031 (0.004) [<0.001]	-0.020 (0.005) [<0.001]	-0.020 (0.005) [<0.001]	-0.025 (0.005) [<0.001]	-0.027 (0.005) [<0.001]	-0.027 (0.005) [<0.001]	-0.026 (0.005) [<0.001]
Additional Controls							
Insurer Fixed Effects		x	x	x	x	x	x
Injury Day-of-the-Week Fixed Effects			x	x	x	x	x
Age Fixed Effects				x	x	x	x
Diagnosis Code Fixed Effects					x	x	x
Indicator for First Medical Treatment at ED						x	x
Medical Spending on First Treatment Date							x
Mean of Dep. Var.	0.619	0.619	0.619	0.619	0.619	0.619	0.619
N	71,366	71,366	71,366	71,366	71,366	71,366	71,366

Notes: This table displays estimates of the coefficients on the female doctor indicator variable, the female claimant indicator variable, and the interaction of the female doctor and female claimant indicator variables from OLS regressions of Equation (1) that control for credential-by-county fixed effects, exam year fixed effects, and injury year fixed effects. Each column represents a separate regression with the dependent variable being I(Additional Cash Benefits > 0). The specifications include additional controls as noted in the table. The diagnosis code fixed effects are based on the first three digits of the ICD-9 code from the earliest treatment. The sample includes claims occurring from 2013 to 2017 that had an independent medical exam by the end of 2017. Standard errors clustered at the doctor level are reported in parentheses, and p-values are reported in brackets.

Table 8: Effect on Amount of Additional Benefits after Controlling for Additional Claimant and Injury Characteristics

	Dependent Variable: Additional Cash Benefits						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Doctor X Female Claimant	0.085 (0.030) [0.005]	0.077 (0.031) [0.013]	0.077 (0.031) [0.013]	0.079 (0.031) [0.010]	0.076 (0.031) [0.012]	0.077 (0.030) [0.012]	0.076 (0.030) [0.013]
Female Doctor	0.031 (0.026) [0.221]	0.032 (0.026) [0.210]	0.032 (0.026) [0.211]	0.030 (0.026) [0.239]	0.030 (0.026) [0.243]	0.029 (0.026) [0.253]	0.030 (0.026) [0.246]
Female Claimant	-0.139 (0.013) [<0.001]	-0.087 (0.014) [<0.001]	-0.087 (0.014) [<0.001]	-0.103 (0.014) [<0.001]	-0.105 (0.014) [<0.001]	-0.104 (0.014) [<0.001]	-0.104 (0.014) [<0.001]
Additional Controls							
Insurer Fixed Effects		x	x	x	x	x	x
Injury Day-of-the-Week Fixed Effects			x	x	x	x	x
Age Fixed Effects				x	x	x	x
Diagnosis Code Fixed Effects					x	x	x
Indicator for First Medical Treatment at ED						x	x
Medical Spending on First Treatment Date							x
Mean of Dep. Var.	6,284	6,284	6,284	6,284	6,284	6,284	6,284
N	71,366	71,366	71,366	71,366	71,366	71,366	71,366

Notes: This table displays estimates of the coefficients on the female doctor indicator variable, the female claimant indicator variable, and the interaction of the female doctor and female claimant indicator variables from Poisson regressions of Equation (1) that control for credential-by-county fixed effects, exam year fixed effects, and injury year fixed effects. Each column represents a separate regression with the dependent variable being the amount of additional normalized benefits received after the exam. The specifications include additional controls as noted in the table. The diagnosis code fixed effects are based on the first three digits of the ICD-9 code from the earliest treatment. The sample includes claims occurring from 2013 to 2017 that had an independent medical exam by the end of 2017. Standard errors clustered at the doctor level are reported in parentheses, and p-values are reported in brackets.

Table 9: Broader Evidence: Gender Differences in Cash Benefit Receipt

	Dependent Variable: I(Cash Benefits > 0)					
	(1)	(2)	(3)	(4)	(5)	(6)
Female Claimant	-0.055 (0.001) [<0.001]	-0.036 (0.001) [<0.001]	-0.033 (0.001) [<0.001]	-0.040 (0.001) [<0.001]	-0.040 (0.001) [<0.001]	-0.039 (0.001) [<0.001]
Controls						
Insurer Fixed Effects		x	x	x	x	x
Injury Month-Year Fixed Effects			x	x	x	x
Injury Day-of-the-Week Fixed Effects			x	x	x	x
Claimant County Fixed Effects			x	x	x	x
Age Fixed Effects				x	x	x
Diagnosis Code Fixed Effects					x	x
Indicator for First Medical Treatment at ED						x
Mean of Dep. Var.	0.215	0.215	0.215	0.215	0.215	0.215
Mean of Dep. Var. for Males	0.236	0.236	0.236	0.236	0.236	0.236
N	1,076,759	1,076,759	1,076,759	1,076,759	1,076,759	1,076,759

Notes: This table displays coefficients on a female claimant indicator variable from separate OLS regressions of Equation (4). The specifications include controls as noted in the table. The diagnosis code fixed effects are based on the first three digits of the ICD-9 code from the earliest treatment. The sample contains 1,076,759 claims occurring from 2013 to 2017. Standard errors are reported in parentheses, and p-values are reported in brackets.

Table 10: Broader Evidence: Gender Homophily in Choice of Providers

	Dependent Variable: I(Chosen "Treating Doctor" is Female)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Claimant	0.023 (0.001) [<0.001]	0.018 (0.001) [<0.001]	0.017 (0.001) [<0.001]	0.016 (0.001) [<0.001]	0.016 (0.001) [<0.001]	0.017 (0.001) [<0.001]	0.016 (0.002) [<0.001]
Controls							
Patient Medical Market (HSA) Fixed Effects		x	x	x	x	x	x
Insurer Fixed Effects			x	x	x	x	x
Injury Month-Year Fixed Effects				x	x	x	x
Diagnosis Code Fixed Effects					x	x	x
Age Fixed Effects						x	x
Patient Medical Market (HSA) X Injury Month-Year							
X Diagnosis Code Fixed Effects							x
Mean of Dep. Var.	0.284	0.284	0.284	0.284	0.284	0.284	0.284
Mean of Dep. Var. for Males	0.276	0.276	0.276	0.276	0.276	0.276	0.276
N	518,536	518,536	518,536	518,536	518,536	518,536	518,536

Notes: This table displays coefficients on a female claimant indicator variable from separate OLS regressions of Equation (5). The sample contains the 518,536 claims occurring from 2013 to 2017 for which we are able to identify the claimants' treating doctor and the gender of the treating doctor. The specifications include controls as noted in the table. The diagnosis code fixed effects are based on the first three digits of the ICD-9 code from the earliest treatment. Standard errors are reported in parentheses, and p-values are reported in brackets.

Table 11: Survey: Experiences with the Health Care System

	Female	Male	Difference	p-value	Regression - Female Coefficient		
	(1)	(2)	(3)	(4)	Est	Std Error	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Thinking about your experiences with health care visits in the past, have you ever felt that a doctor did any of the following? - Share answered yes</i>							
Talked down to you or didn't treat you with dignity or respect	0.393	0.340	0.052	0.035	0.087	(0.026)	0.001
Didn't understand or relate to your experiences and concerns	0.450	0.356	0.094	<0.001	0.104	(0.027)	<0.001
Didn't believe you were telling the truth about your symptoms or concerns	0.317	0.286	0.031	0.193	0.045	(0.025)	0.068
Refused to order a test or treatment you thought you needed	0.223	0.232	-0.009	0.666	0.022	(0.022)	0.310
Made you feel uncomfortable discussing your concerns	0.292	0.250	0.042	0.069	0.059	(0.024)	0.014
Assumed something about you without asking	0.357	0.328	0.029	0.236	0.050	(0.026)	0.054
Refused to prescribe pain medication you thought you needed	0.175	0.249	-0.074	<0.001	-0.055	(0.021)	0.008
Share answered yes to any of the above	0.596	0.538	0.058	0.023	0.103	(0.026)	<0.001
<i>Thinking about your experiences getting health care for yourself, which doctor---male or female--- would be more likely to [Answer options: male doctor, female doctor, male and female doctors are equally likely]</i>							
<i>Share selected doctor of own-gender</i>							
treat you with dignity and respect?	0.343	0.193	0.150	<0.001	0.178	(0.024)	<0.001
understand or relate to your experiences and concerns?	0.517	0.295	0.222	<0.001	0.232	(0.027)	<0.001
believe you are telling the truth about your symptoms or concerns?	0.304	0.175	0.129	<0.001	0.159	(0.023)	<0.001
provide needed testing and treatments?	0.249	0.176	0.073	0.001	0.103	(0.022)	<0.001
make you feel comfortable with discussing your concerns?	0.416	0.240	0.176	<0.001	0.179	(0.026)	<0.001
ask appropriate questions instead of making assumptions?	0.326	0.177	0.148	<0.001	0.177	(0.023)	<0.001
be the most qualified?	0.161	0.160	0.001	0.969	0.016	(0.020)	0.406
be available near you?	0.154	0.184	-0.030	0.124	-0.007	(0.020)	0.733
<i>Share did not select doctor of opposite gender</i>							
treat you with dignity and respect?	0.931	0.820	0.111	<0.001	0.108	(0.018)	<0.001
understand or relate to your experiences and concerns?	0.932	0.826	0.105	<0.001	0.088	(0.018)	<0.001
believe you are telling the truth about your symptoms or concerns?	0.928	0.816	0.112	<0.001	0.114	(0.018)	<0.001
provide needed testing and treatments?	0.932	0.853	0.079	<0.001	0.079	(0.017)	<0.001
make you feel comfortable with discussing your concerns?	0.932	0.837	0.095	<0.001	0.090	(0.017)	<0.001
ask appropriate questions instead of making assumptions?	0.933	0.813	0.120	<0.001	0.120	(0.018)	<0.001
be the most qualified?	0.958	0.890	0.068	<0.001	0.053	(0.014)	<0.001
be available near you?	0.914	0.898	0.016	0.298	0.011	(0.016)	0.501

Notes: This table summarizes responses from several questions from the survey described in Section 7. The table above reports the raw means for each of the indicated variables, along with the difference by respondent gender and the associated p-value testing whether the difference is zero. The table also includes the estimated coefficient on female (and associated standard error and p-value) from a regression of the indicated response on a female indicator and all the respondent characteristics reported in Appendix Table A3 (as in Equation 6). The survey was conducted by Qualtrics and included 1,519 adults ages 30 to 64. Eligibility was restricted to individuals who reported working at some point in the last 12 months and individuals who self-identified as either male (755 respondents) or female (764 respondents).

Table 12: Survey: Preferences over Providers and Homophily

	Female (1)	Male (2)	Difference (3)	p-value (4)	Regression - Female Coefficient		
					Est (5)	Std Error (6)	p-value (7)
<i>Have you ever received care from a.... - Share answered yes</i>							
Male doctor?	0.901	0.968	-0.068	<0.001	-0.062	0.013	<0.001
Female doctor?	0.941	0.767	0.174	<0.001	0.162	0.019	<0.001
<i>Given the choice, would you prefer to see a doctor who is male or female, or does it not make much difference to you? - Share select each option below</i>							
Prefer to see a doctor who is female	0.463	0.136	0.327	<0.001	0.315	0.024	<0.001
Prefer to see a doctor who is male	0.058	0.290	-0.232	<0.001	-0.207	0.020	<0.001
Doesn't make much difference	0.479	0.574	-0.094	<0.001	-0.108	0.028	<0.001
<i>Choices based on initial hypothetical choice question: doctor gender and out-of-pocket cost - Share select own-gender doctor</i>							
Own Gender \$30 vs. Opp Gender \$5	0.254	0.240	0.014	0.748	0.029	0.048	0.544
Own Gender \$10 vs. Opp Gender \$5	0.485	0.293	0.192	<0.001	0.200	0.054	<0.001
Own Gender \$5 vs. Opp Gender \$10	0.929	0.833	0.095	0.005	0.067	0.036	0.064
Own Gender \$5 vs. Opp Gender \$30	0.988	0.854	0.134	<0.001	0.129	0.031	<0.001
<i>Choices based on full set of hypothetical choice questions: doctor gender and out-of-pocket cost - Share select own-gender doctor</i>							
Own Gender \$30 vs. Opp Gender \$5	0.229	0.147	0.082	<0.001	0.091	0.021	<0.001
Own Gender \$10 vs. Opp Gender \$5	0.427	0.273	0.154	<0.001	0.159	0.026	<0.001
Own Gender \$5 vs. Opp Gender \$10	0.929	0.828	0.102	<0.001	0.088	0.018	<0.001
Own Gender \$5 vs. Opp Gender \$30	0.970	0.910	0.060	<0.001	0.050	0.014	<0.001
<i>Please indicate how important each of the following characteristics is if you were choosing a doctor - Share indicating at least moderately important</i>							
Out-of-pocket cost for a visit	0.829	0.824	0.005	0.809	0.005	0.022	0.807
Doctor reviews (e.g., on websites like Healthgrades or Google or from friends or family)	0.847	0.807	0.040	0.038	0.048	0.022	0.027
Travel time to get to doctor's office	0.831	0.834	-0.003	0.864	0.002	0.021	0.914
Wait time at the doctor's office	0.863	0.848	0.015	0.410	0.018	0.020	0.361
Doctor's sex	0.414	0.336	0.077	0.002	0.112	0.026	<0.001
Doctor's age	0.349	0.370	-0.020	0.416	0.021	0.026	0.415

Notes: This table summarizes responses from several questions from the survey described in Section 7. The table above reports the raw means for each of the indicated variables, along with the difference by respondent gender and the associated p-value testing whether the difference is zero. The table also includes the estimated coefficient on female (and associated standard error and p-value) from a regression of the indicated response on a female indicator and all the respondent characteristics reported in Appendix Table A3 (as in Equation 6). The survey was conducted by Qualtrics and included 1,519 adults ages 30 to 64. Eligibility was restricted to individuals who reported working at some point in the last 12 months and individuals who self-identified as either male (755 respondents) or female (764 respondents).

APPENDIX

A More Detail on Policy Counterfactuals

The goal of this analysis is to understand how gender gaps conditional on observables are impacted by: (i) the share of treating doctors who are female (r) and (ii) the degree of gender homophily in patient-provider matches (Z). In other words, we would like to express the gender gap (conditional on observables) as a function of the overall share of treating doctors who are female and the degree of homophily in the market: $Gap(r, Z)$. We derive this function in two steps. First, following directly from the definitions in the text, we can express the share of female claimants seeing female doctors (r_f) as a function of the share of female treating doctors overall (r) and the degree of relative gender homophily (Z):

$$r_f = \frac{r}{\frac{1-s}{Z} + s}. \quad (1)$$

Second, we can use our estimate of the gender-match effect (X) and the definitions to infer the values of Gap_m and Gap_f :

$$Gap_m \equiv \frac{Gap^0}{r_f^0(X-1) + 1} \quad (2)$$

$$Gap_f \equiv \frac{XGap^0}{r_f^0(X-1) + 1}, \quad (3)$$

where Gap^0 and r_f^0 , respectively, represent the observed values of the gender gap and the share of female patients selecting female treating doctors in the status quo.

Putting these together, we obtain:

$$Gap(r, Z) = Gap_m \times \left(1 - \frac{r}{\frac{1-s}{Z} + s}\right) + Gap_f \times \left(\frac{r}{\frac{1-s}{Z} + s}\right), \quad (4)$$

where $Gap_f \equiv \frac{XGap^0}{r_f^0(X-1)+1}$ and $Gap_m \equiv \frac{Gap^0}{r_f^0(X-1)+1}$.

Table A1: Effect on Indicator for Receiving Additional Benefits after Controlling for Doctor Experience

	Dependent Variable: I(Additional Cash Benefits > 0)		
	(1)	(2)	(3)
Female Doctor X Female Claimant	0.030 (0.010) [0.003]	0.031 (0.010) [0.002]	0.031 (0.010) [0.002]
Female Doctor	-0.0002 (0.0080) [0.984]	-0.00002 (0.00801) [0.998]	-0.00001 (0.00802) [0.999]
Female Claimant	-0.031 (0.004) [<0.001]	-0.035 (0.007) [<0.001]	-0.043 (0.014) [0.002]
Doctor's Number of Exams in Previous Year		0.00005 (0.00011) [0.640]	0.0001 (0.0001) [0.594]
Doctor's Number of Exams in Previous Year X Female Claimant		0.0001 (0.0001) [0.458]	0.0001 (0.0001) [0.551]
Doctor's Share of Exams in Previous Year for Females			-0.012 (0.022) [0.588]
Doctor's Share of Exams in Previous Year for Females X Female Claimant			0.027 (0.032) [0.403]
Indicator for Doctor Performing No Exams in Previous Year			0.021 (0.020) [0.297]
Indicator for Doctor Performing No Exams in Previous Year X Female Claimant			-0.050 (0.036) [0.169]
Mean of Dep. Var.	0.619	0.619	0.619
N	71,366	71,366	71,366

Notes: This table displays estimates of the coefficients on the female doctor indicator variable, the female claimant indicator variable, and the interaction of the female doctor and female claimant indicator variables from OLS regressions of Equation (1) that control for credential-by-county fixed effects, exam year fixed effects, and injury year fixed effects. Each column represents a separate regression with the dependent variable being I(Additional Cash Benefits > 0). The specifications include additional controls as noted in the table. For 987 exams, doctors had not performed an exam in the previous year. We set doctors' share of exams for females in the previous year to zero for these exams and include an indicator variable for the doctor not having performed an exam in the previous year in addition to the control for the share of exams for females. The sample includes claims occurring from 2013 to 2017 that had an independent medical exam by the end of 2017. Standard errors clustered at the doctor level are reported in parentheses, and p-values are reported in brackets.

Table A2: Robustness to Varying Fixed Effects

	Dependent Variable: I(Additional Cash Benefits > 0)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Doctor X Female Claimant	0.030 (0.010) [0.003]	0.032 (0.010) [0.002]	0.039 (0.014) [0.006]	0.029 (0.010) [0.005]	0.029 (0.010) [0.005]	0.034 (0.011) [0.002]	0.028 (0.010) [0.008]
Female Doctor	-0.000 (0.008) [0.984]	-0.001 (0.008) [0.929]		0.000 (0.008) [0.955]	-0.000 (0.008) [0.987]	-0.001 (0.008) [0.885]	0.000 (0.008) [0.986]
Female Claimant	-0.031 (0.004) [<0.001]	-0.031 (0.005) [<0.001]	-0.019 (0.006) [0.002]			-0.035 (0.005) [<0.001]	-0.021 (0.005) [<0.001]
Additional Controls							
County by Credential Fixed Effects	x			x	x		x
Exam Year Fixed Effects	x					x	
County by Credential by Exam Year Fixed Effects		x					
County by Provider by Exam Year Fixed Effects			x				
Gender by County and Gender by Exam Year Fixed Effects				x			
Gender by County by Exam Year Fixed Effects					x		
County by Diagnosis Fixed Effects						x	
Insurer by Exam Year Fixed Effects							x
Mean of Dep. Var.	0.619	0.619	0.619	0.619	0.619	0.619	0.619
N	71,366	71,366	71,366	71,366	71,366	71,366	71,366

Notes: This table displays estimates of the coefficients on the female doctor indicator variable, the female claimant indicator variable, and the interaction of the female doctor and female claimant indicator variables. Each column represents a separate regression with the dependent variable being I(Additional Cash Benefits > 0). In addition to the controls indicated in the table, each specification also controls for claimants' injury year. The sample includes claims occurring from 2013 to 2017 that had an independent medical exam by the end of 2017. Standard errors clustered at the doctor level are reported in parentheses, and p-values are reported in brackets.

Table A3: Survey: Respondent Characteristics

	Survey Respondents				Workers in ACS		Workers' Comp Claimants	
	Female (1)	Male (2)	Difference	p-value	Female (3)	Male (4)	Female (5)	Male (6)
Ever unemployed in last 12 months	0.270	0.225	0.044	0.045	-	-	-	-
Married	0.516	0.640	-0.124	<0.001	0.590	0.639	0.454	0.558
Hispanic	0.093	0.074	0.019	0.187	0.158	0.184	-	-
Works in healthcare	0.179	0.117	0.063	0.001	0.193	0.052	-	-
Has health insurance	0.878	0.891	-0.013	0.424	0.908	0.867	-	-
Interacted with a doctor to get health care in last five years	0.932	0.918	0.014	0.299	-	-	-	-
Experienced chronic physical pain that interfered with daily activities in last five years	0.442	0.491	-0.049	0.056	-	-	-	-
Age:								
30-39 years	0.355	0.270	0.085	<0.001	0.314	0.322	0.262	0.339
40-49 years	0.260	0.351	-0.091	<0.001	0.292	0.292	0.315	0.311
50-55 years	0.296	0.309	-0.013	0.587	0.282	0.276	0.296	0.244
60-64 years	0.089	0.070	0.019	0.176	0.112	0.110	0.092	0.076
Race:								
White	0.822	0.837	-0.015	0.434	0.721	0.744	-	-
Black or African American	0.085	0.073	0.012	0.377	0.135	0.106	-	-
Asian	0.043	0.036	0.007	0.457	0.069	0.066	-	-
Other	0.038	0.045	-0.007	0.490	0.074	0.085	-	-
Decline to State	0.012	0.009	0.003	0.632	-	-	-	-
Highest Level of Education:								
< High School	0.009	0.011	-0.001	0.778	0.068	0.103	-	-
High School Graduate or GED	0.181	0.146	0.035	0.065	0.208	0.266	-	-
Some College	0.344	0.225	0.119	<0.001	0.309	0.277	-	-
Bachelor's Degree	0.291	0.350	-0.059	0.014	0.245	0.218	-	-
Postgraduate Degree	0.175	0.269	-0.093	<0.001	0.171	0.136	-	-
Annual Earnings:								
<10K	0.064	0.029	0.035	0.001	0.103	0.054	0.030	0.010
10-20K	0.084	0.034	0.049	<0.001	0.129	0.071	0.168	0.054
20-30K	0.134	0.056	0.078	<0.001	0.146	0.103	0.292	0.155
30-40K	0.140	0.090	0.050	0.002	0.139	0.119	0.166	0.180
40-50K	0.115	0.095	0.020	0.208	0.109	0.109	0.104	0.163
50-75K	0.208	0.179	0.029	0.148	0.186	0.215	0.140	0.263
75-100K	0.135	0.225	-0.090	<0.001	0.086	0.117	0.073	0.117
>100K	0.120	0.291	-0.171	<0.001	0.102	0.213	0.027	0.058
Industry:								
Agriculture/Forestry/Fishing/Hunting	0.014	0.026	-0.012	0.096	0.007	0.018	0.006	0.014
Arts/Entertainment/Accommodation/Food Services	0.064	0.061	0.003	0.796	0.075	0.065	0.065	0.033
Information/Finance/Real Estate/Professional Services	0.170	0.274	-0.104	<0.001	0.208	0.217	0.071	0.065
Health Care/Educational Services	0.297	0.131	0.166	<0.001	0.376	0.113	0.432	0.086
Manufacturing	0.063	0.136	-0.074	<0.001	0.068	0.145	0.069	0.147
Mining/Utilities/Construction	0.037	0.077	-0.040	0.001	0.023	0.153	0.018	0.198
Public Administration/Other Services	0.190	0.142	0.048	0.012	0.103	0.094	0.193	0.259
Wholesale Trade/Retail Trade/Transportation	0.165	0.152	0.013	0.502	0.140	0.196	0.146	0.199
Region:								
West	0.174	0.208	-0.034	0.093	0.232	0.246	-	-
Midwest	0.246	0.197	0.049	0.022	0.212	0.210	-	-
Northeast	0.212	0.217	-0.005	0.806	0.181	0.176	-	-
South	0.368	0.377	-0.010	0.697	0.375	0.369	-	-

Notes: Columns 1 and 2 of this table display respondent characteristics from the survey described in Section 7. The survey was conducted by Qualtrics and included 1,519 adults ages 30 to 64. Eligibility was restricted to individuals who reported working at some point in the last 12 months and individuals who self-identified as either male (755 respondents) or female (764 respondents). For comparison, columns 3 through 6 of the table also displays characteristics of workers ages 30 to 64 from the 2019 American Community Survey and characteristics of workers' compensation claimants in Texas ages 30 to 64 injured between 2013 and 2017. As described in Section 3, the presence of wage and industry information is related to the receipt of cash benefits in the workers' compensation data. All wages are in 2020 dollars, and the American Community Survey numbers are weighted using IPUMS weights.