

Interpreting Signals in the Labor Market: Evidence from Medical Referrals

Heather Sarsons*

November 28, 2017

[Click here for most recent version](#)

Abstract

This paper provides evidence that a person's gender influences the way others interpret information about his or her ability and documents the implications for gender inequality in labor markets. Using data on physicians' referrals to surgical specialists, I find that the referring physician views patient outcomes differently depending on the performing surgeon's gender. Physicians become more pessimistic about a female surgeon's ability than a male's after a patient death, indicated by a sharper drop in referrals to the female surgeon. However, physicians become more optimistic about a male surgeon's ability after a good patient outcome, indicated by a larger increase in the number of referrals the male surgeon receives. After a bad experience with one female surgeon, physicians also become less likely to refer to new female surgeons in the same specialty. There are no such spillovers to other men after a bad experience with one male surgeon. Consistent with learning models, physicians' reactions to events are strongest when they are beginning to refer to a surgeon. However, the empirical patterns are only consistent with Bayesian learning if physicians do not have rational expectations about the true distribution of surgeon ability.

*sarsons@fas.harvard.edu

I am indebted to my advisors, Claudia Goldin, Lawrence Katz, David Laibson, and Amanda Pallais, for their guidance and encouragement. I am also grateful to David Cutler for his help obtaining and understanding the Medicare data, and to my informal advisor, Ben Enke, for looking over numerous early drafts. I thank Mitra Akhtari, Kai Barron, Anke Becker, Kirill Borusyak, Amitabh Chandra, Sylvain Chassang, John Coghlanese, Oren Danieli, Krishna Dasartha, Ellora Derenoncourt, Raissa Fabregas, John Friedman, Roland Fryer, Peter Ganong, Edward Glaeser, Joshua Goodman, Nathan Hendren, Simon Jäger, Anupam Jena, Jessica Laird, Jing Li, Jonathan Libgober, Matthew Rabin, Hannah Schafer, Frank Schilbach, Nihar Shah, Andrei Shleifer, Jann Spiess, Neil Thakral, Peter Tu, Shosh Vasserman, and Chenzi Xu for helpful comments. This paper also benefitted from comments from seminars at Harvard, the Wharton School, Carnegie Mellon, and the ECBE conference. Dr. Nancy Keating, Dr. Erika Rangel, and several other unnamed physicians provided valuable institutional knowledge. Mohan Ramanujan and Jean Roth provided valuable data support.

1 Introduction

Does a person's gender influence the way we interpret information about his or her ability? The answer to this question has important implications for gender inequality in labor markets, particularly for how women move up the career ladder relative to men. Research shows that in many industries, women are promoted at lower rates than their male counterparts.¹ Gender gaps in wages persist in upper-level positions (Blau and Kahn, 2016). Employers use information that is often subjective or imprecise to evaluate individuals. Their evaluations influence hiring, wage, and promotion decisions. Such decisions become distorted if there are systematic differences in how information about a man and a woman is processed. If an employee's gender influences the way an employer views his or her performance, for example, the employer might end up with different beliefs about a man and a woman's ability even if their objective performance is the same.

This paper empirically tests whether gender influences the way information about others is interpreted. Using Medicare data on referrals from physicians to surgical specialists, I find that the referring physicians view their patients' surgical outcomes differently depending on the performing surgeon's gender.² Physicians increase their referrals more to a male surgeon than to a female surgeon after a good patient outcome but lower their referrals more to a female surgeon than a male surgeon after a bad outcome. Furthermore, a physician's experience with one female surgeon influences his or her referrals to other female surgeons in the same specialty. An experience with a male surgeon has no impact on a physician's behavior toward other male surgeons. These asymmetric responses imply that even if women are hired at the same rate as men, they receive fewer chances to show that they can be successful which could lead to lower promotion rates and wages.

Three features of the medical field and Medicare data allow me to address whether gender influences belief updating. First, medical research suggests that a primary factor influencing physicians' referral choices is their beliefs about a surgeon's ability (Barnett et al., 2007; Forrest et al., 2006; and Kinchen et al., 2004). The volume of a physician's referrals to a surgeon thus provides a proxy for the physician's belief about that surgeon's ability. Second, the high frequency of referral decisions makes it possible to document how

¹For example, women represent half of all managers in Fortune 500 companies but only 14% of all executive officers and 4% of all CEOs (Blau and Kahn, 2016). Women are less likely to make partner within law firms (Azmat and Ferrer; 2017; Noonan, Corcoran, and Courant; 2005). More than 40% of all medical students are female but only 24% of hospital division chiefs and 15% of medical department chairs are women (Lautenberger et al., 2014). Within academia, women are less likely to move to associate and full professorship positions than men (Blau and Kahn, 2016).

²Throughout the paper, I use the term physician for the referring physician and surgeon for the surgeon who performs the surgery even though surgeons are also physicians.

physicians react to individual signals, something that is difficult to do in many work contexts.³ Finally, using detailed Medicare data allows me to control for variables that could influence a physician's reaction, such as the patient and procedure risk, or the surgeon's experience. I can thereby isolate the portion of a physician's reaction that is attributable to the surgeon's gender.

To study the influence that gender has on a physician's reaction, I identify and match observably similar male and female surgeons who perform the same procedure on similar patients. Performing an event study on this matched sample, I document how the referring physician's belief about the performing surgeon, as well as other surgeons, changes after good and bad outcomes. The analysis reveals two asymmetries.

First, physicians update asymmetrically about individual male and female surgeons. Following a bad patient outcome (a patient death), physicians lower their beliefs about a female surgeon's ability more than they do for male surgeons. Referrals from the physician drop by 34% after a bad outcome when the surgeon is female compared with only a slight stagnation in referrals when the surgeon is male. Following a good outcome (an unanticipated survival), however, physicians become more optimistic about a male surgeon's ability than a female surgeon's, indicated by a doubling of referrals to the man versus a 70% increase in referrals to the woman

The second asymmetry exists in how physicians treat groups. Physicians appear to use information about individual female surgeons to update their beliefs about other female surgeons in the same specialty. Physicians become less likely to form new referral connections with women after a bad experience with one female surgeon. In contrast, a bad experience with one male surgeon does not affect physicians' behavior toward other men. I find weak evidence of positive spillovers to other women after a physician has a good experience with a female surgeon.

Two pieces of evidence suggest that the drop in referrals to women following a bad outcome is largely driven by the physician's behavior rather than female surgeons turning down referrals. First, the fact that physicians become less likely to refer to other women in the same specialty following a death suggests that physicians change their beliefs about female surgeons. Second, female surgeons do not receive fewer referrals from other physicians following a patient death, suggesting that they do not generally become more risk averse and refuse to perform surgeries in the future.

Looking at which physicians update asymmetrically, I find that the effects are concentrated among physicians who just started referring to a particular surgeon. Physicians

³Promotion decisions and pay raises, for example, are viewed only after employers have received a series of signals about an employee, making it difficult to discern how they interpret each signal.

both react less strongly to signals and treat men and women more equally the longer is their referral history with a particular surgeon. Physicians thus appear to learn about surgeon ability over time and exhibit asymmetric learning only when they first start referring to a surgeon. Physicians who have more experience working with female surgeons are also more likely to treat men and women equally in response to an outcome. The physician's gender does not seem to play a role in how he or she reacts.

Examining the career implications of asymmetric updating, I find that in addition to receiving fewer referrals, women also receive less difficult procedures and less risky patients after a patient death. This change in the types of referrals female surgeons receive affects both skill accumulation and surgeon pay. Since surgeons learn by performing surgeries, women will have fewer chances to develop their surgical skills as they receive fewer referrals for difficult surgeries. Procedure risk is also correlated with pay. Overall, I find that women lose 60% of their Medicare billings from the referring physician per quarter when they experience a bad patient outcome, whereas men lose 30%.

Whether changes in referrals truly reflect differences in how physicians process signals depends on whether alternative interpretations can be ruled out. I consider three main alternative explanations and show that none is supported in the data. I first test whether male surgeons systematically receive riskier patients than women. If they do, physicians would naturally react less to a death under a male surgeon than a female surgeon. Although I match on observed patient risk, unobservable risk differences could exist. I bound what the difference in unobservable risk between men and women's patients would have to be to justify the physician's reaction. I find that men would have to receive patients who are 70 percentage points riskier on unobservables for risk to explain the gender difference in a physician's reaction.

A second alternative is that patient deaths are more predictive of future deaths for women while unanticipated survivals are more predictive of future survivals for men. In this case, physicians will become more certain that a woman is low ability after a death and more certain that a man is high ability after a survival, generating the asymmetry observed in the data. I check for differential predictability of future events conditional on a surgeon having one such event and conditional on the characteristics of past and future patients. I find that although a death or survival is predictive of a future death or survival, there is no difference by surgeon gender. In fact, women are slightly less likely to have another bad outcome after one patient dies.

Finally, physicians could become risk averse after a patient death and stop referring for a particular procedure. If women specialize in those procedures whereas men perform a wider variety of procedures, it will look like referrals to women fall after a death when

in fact the physician stops referring to all surgeons for a particular procedure. I look at how the types of procedures that physicians refer for change after a patient death and do not find evidence of any behavior change.

Turning to what the empirical results tell us about how physicians update, I draw on Zeltzer (2017) to develop a model of referrals decisions in which physicians try to maximize patient outcomes subject to idiosyncratic factors like patient preferences and wait times. I use the model to understand whether the empirical results are in line with Bayesian updating, considering cases in which physicians do and do not have risk preferences.

I present two propositions that place restrictions on what a physician's priors about men and women's abilities would have to look like for the results to be consistent with Bayesian updating. I show that a physician would either have to believe that women are on average higher ability than men, or that the difference in the variance of women and men's abilities increases as a physician receives more signals. I discuss whether these assumptions are plausible given the data and argue that although physicians could be Bayesian, their beliefs about ability would not be in line with men and women's measured ability distributions, ruling out rational expectations.

I then consider alternative models, first discussing how the empirical results are at odds with existing discrimination models, and then describing a model of attribution bias. In this model, physicians have an expected outcome for each surgery. When the actual outcome matches their expected outcome, physicians update as Bayesians. However, when the actual outcome is far from what they expected, physicians rationalize the event by attributing it either to the physician's ability or to noise, relying on existing stereotypes about men and women to do so. Therefore, if there is a broad stereotype that women are worse surgeons than men, physicians will attribute unexpectedly bad events to a woman's ability and unexpectedly good events to noise.

The model also explains the asymmetry in spillovers to other male and female surgeons. Physicians update iteratively about the group upon seeing outcomes from individuals from that group. Because women are underrepresented as surgeons, physicians receive fewer signals from them relative to men. As in a Bayesian model, each signal has a larger impact on how a physician updates about the female ability distribution than it does for men. In contrast to a Bayesian model, though, physicians update too little about the group in the positive direction as they attribute some of women's good signals to noise. The asymmetry in the attribution of signals from individual surgeons generates the empirical observation that bad outcomes spill over to other female surgeons while good outcomes have only a weak effect on other female surgeons.

This paper relates to several literatures. First, a large body of work has sought to explain the gender gap in men and women's labor market outcomes. Factors such as family commitments, work preferences, personality traits, and discrimination have all been shown to contribute to the gap.⁴ Yet a large portion remains unexplained. A growing literature also documents a pay gap between male and female physicians. Jena et al. (2016), for example, find that female physicians make \$19,878 less per year than male physicians after adjusting for specialty and a variety of physician characteristics. Among physicians accepting Medicare, Zeltzer (2017) documents significant gender homophily in referral decisions, leading to 5% lower demand for female physicians than for male physicians. This paper provides another mechanism that might explain part of the gender gap in wages and promotions, particularly among surgeons.

It also relates to the large literature on discrimination. Most papers in this area involve employers discriminating at the beginning of someone's career and then learning about a worker's true ability over time.⁵ In these papers, employers have rational expectations about worker effort or ability. This paper offers a new angle to this discussion. Even if two workers are treated similarly early in their careers, employers might treat them unequally over time if they do not have rational expectations or if they discriminate in how they interpret signals. I cannot distinguish between the two, but neither fit with the current discrimination literature.

There is also empirical evidence that people from different demographic groups are punished differently for the same behavior or outcome. Female financial advisors, for example, are more likely to be fired for misconduct than are male advisors (Egan et al., 2017). Black students are more likely than white students to be suspended for misbehaving in class (Skiba et al., 2002; Gregory et al., 2010). Less studied, though, is whether there is differential treatment for good outcomes and whether these differences are driven by differences in individuals' behavior or differences in how others view their behavior. My results suggest that at part of the differential treatment is driven by the individual interpreting the event.

Finally, the paper relates to the behavioral and social psychology literature on biased updating. Asymmetric updating has been documented in the lab when individuals receive signals about themselves. Women are especially prone to updating too little when they receive good signals about their ability and updating too much when they receive

⁴For literature on each topic, see Antecol et al. (2016), Babcock and Laschever (2004), Blau and Kahn (2016), Bursztyn et al. (2017), Card et al. (2016), Ceci et al. (2014), Ginther and Kahn (2004), Goldin (2014a), Goldin (2014b), Goldin and Rouse (2000), Kleven et al. (2017), and Niederle and Vesterlund (2007)

⁵See, for example, Becker (1957), Phelps (1972), Arrow (1973), Coate and Loury (1993), Fryer (2007), Altonji and Pierret (2001), and Bertrand and Mullainathan (2004).

bad signals. Eil and Rao (2011), for example, find that individuals avoid new information about themselves if it is negative but update using Bayes' Rule when they receive positive feedback. Mobius et al. (2014) also provide experimental evidence that individuals overweight good signals and under-weight bad signals about themselves but that women are more conservative when doing so. This paper shows a similar result in a non-lab context and when updating about others rather than oneself.

The remainder of the paper is organized as follows. Section 2 discusses the empirical setting, identifying the factors that influence physicians' referral choices, and the Medicare data. The empirical strategy is outlined in Section 3 and the main results are presented in Section 4. Section 5 explores alternative interpretations of the results. I analyze the impact that asymmetric updating has on surgeon quality and on surgeons' career trajectories in Section 6. In Section 7, I present a theoretical framework to help interpret the empirical results. Section 8 concludes.

2 Empirical Setting and Data

I attempt to overcome the challenge of empirically measuring and tracking beliefs using Medicare data on referrals from physicians to surgeons. The approach relies on the assumption that a physician's referral choices are a valid measure of her beliefs about a surgeon's ability. In this section, I discuss the literature documenting how physicians make their referral decisions before describing the Medicare data.

2.1 Referral Decisions

Several medical studies have attempted to identify the factors influencing doctors' referral decisions. I focus specifically on the literature discussing how decisions about surgical referrals are made as I focus on surgeries in my analysis. While many factors influence a physician's referral choice, these studies, combined with informal interviews, suggest that surgeon quality is a primary consideration in a physician's choice. In fact, several studies ask physicians to rank the reasons for referring a patient to a specific specialist, *aside* from clinical expertise, suggesting that this is an obvious factor⁶. In a survey of 1200 physicians across the U.S., Kinchen et al. (2004) find that 88% of respondents considered the surgeon's skill (measured by medical skill and board certification) to be one of the most important factors in deciding whether to make a referral. Choudhry et al. (2014) also find that the perceived expertise of a specialist is a primary driver of referrals.

⁶See, for example, Barnett et al., 2012a; Barnett et al., 2012b

However, there are several other factors that influence referrals, many of which are not observed in my data and will add noise to my estimates. Surgeon availability and the urgency of the surgery, for example, are important determinants of surgeon choice (Forrest et al., 2006). Patients in worse health are significantly more likely to be referred to the first available surgeon (Shea et al., 1999). Physicians also look for surgeons with good communication skills so that they can easily follow up to manage the patient’s post-surgical recovery (Kinchen et al., 2004; Forrest et al., 2006). Insurance status plays a role in a physician’s referral choice, but the patients I look at all have Medicare coverage, making this factor less relevant. Surprisingly, risk attitudes and tolerance for uncertainty appear to be weak predictors of whom a physician will refer to (Forrest et al., 2006). While these variables do influence the physician’s decision, they would have to be systematically correlated with surgeon gender to explain the results.

Patients can also request certain surgeons and these cases are not always identifiable in the data. In a paper studying how patients choose surgeons, however, Freedman et al. (2015) find that the most frequently reported reason a patient gives for choosing a particular surgeon is that their physician recommended that surgeon. In Section 4, I show that while the results could be partly driven by patient preferences, a large part is due to physicians changing their beliefs and referral behavior.

2.2 Data

The primary data source for this study is the Medicare Carrier file, a 20% random sample of fee-for-service claims of all Medicare beneficiaries in the U.S. between 2008 and 2012. Three features of the Medicare data make it ideal for understanding belief updating. First, the dataset includes detailed information on patients, including diagnoses, demographic information, medical history, and procedure codes. I can therefore compare surgeons who are performing the same procedure on patients with similar demographic backgrounds, medical histories, and risk levels. Second, referrals are frequent, allowing me to document how a physician changes her behavior immediately following a surgery. Finally, the Medicare data lets me track surgeons over time and see how well they perform on surgeries that the referring physician might not witness. I can then calculate the career implications of biased updating and assess whether a physician’s beliefs about a particular surgeon match the surgeon’s actual skill.

I supplement the Carrier file with two other datasets: the Physician Compare National file and the Dartmouth Atlas of Health Care. The Physician Compare National file contains information on physicians and surgeons, such as the doctor’s gender, specialty,

medical school, and experience. It also has information on the hospital or group practice to which the doctor belongs. The Dartmouth Atlas of Health Care is a geographic dataset that lets me to match physicians and surgeons to their Hospital Referral Region (HRR). HRRs are geographic units representing regional health care markets and are the geographic unit within which physicians typically refer⁷. There are 306 HRRs in the U.S. The Atlas also has information on the number of specialists within each HRR so that I know how many surgeons a physician has the option of referring to.

I restrict the data in four ways. I first limit the sample to surgical procedures and specialties, as surgical procedures have clear outcomes, such as a patient death. Second, I only include physicians who have the option of referring to at least two specialists for the procedure they are referring for to ensure that their behavior is not constrained by the number of specialists in a region. I then limit the sample to cases in which one surgeon performed a procedure, rather than a team of surgeons. Although there would still be others in the operating room, such as an anesthesiologist and nurses, restricting to cases where there is a single lead surgeon means that, as much as possible, there is a clearly identified person responsible for the case. Finally, I include physician-surgeon pairs for which there was at least one referral prior to the patient event so that I can observe pre-trends.

2.2.1 Primary Variable Construction

Referrals Referrals between a physician and a surgeon are identifiable in the Medicare data. For each patient diagnosis, I see whether the patient was referred to a surgeon. I use the term physician when describing any doctor who makes a referral to a surgeon. These physicians can be primary care doctors or specialists.

I do not see instances in which a physician refers a patient to a surgeon and the surgeon turns down the patient. As such, referrals are defined as those that were actually followed through: a physician refers a patient to a surgeon and the surgeon sees the patient. The results could therefore be driven by the surgeon's behavior if women are more likely to stop taking referrals after a bad outcome. However, in Section 4, I show that the results are not explained by surgeons changing their behavior. I exclude referrals from medical professionals other than doctors, such as nurse practitioners, as well as self-referrals in which a specialist refers a patient to him or herself.⁸

⁷For more information, see <http://www.dartmouthatlas.org/data/region/>.

⁸Self-referrals are often cases where a patient requests a particular surgeon.

Patient Risk If women receive less risky patients than men, it would be natural for a physician to react more strongly to a patient death under a woman. I therefore construct and match on a measure of patient risk to ensure that I am comparing male and female surgeons who are receiving similar patients. I follow the medical literature to calculate patient risk, combining a comorbidity index with patient characteristics to predict patient mortality for a given procedure⁹. Specifically, I first use ICD-9 diagnosis codes to calculate the Elixhauser Index for each patient visit.¹⁰ The Elixhauser Index uses information on patients’ medical histories to categorize comorbidities, pre-existing medical conditions known to increase the risk of death. I then use the Elixhauser Index, along with other patient characteristics such as age, gender, and race, to predict in-hospital mortality. The higher is the probability of death, the riskier is the patient.

Surgeon Ability I construct a measure of surgeon ability for two reasons. First, I use it to estimate the true distribution of surgeon ability and compare it to what a Bayesian physician’s prior distribution over ability would have to look like to explain the empirical results. Second, I track how the average ability of surgeons that a physician refers to changes when they switch surgeons.

I calculate surgeon ability as follows. For each patient a surgeon sees, I take the difference between the patient’s risk of death for the procedure being performed (defined above) and an indicator variable that equals one if the patient died. I then take the average over all of the patients that a surgeon sees:

$$Ability_i = \frac{\sum_{p_i} (Risk_p - \mathbb{1}(Death))}{n_i} \quad (1)$$

Here, n_i is the number of patients that surgeon i sees, $Risk_p$ is patient p ’s risk of death for the procedure being performed, and $\mathbb{1}(Death)$ is an indicator if patient p died.

Events (Signals) I consider two signals that surgeons send to physicians: bad and good. Bad signals are defined as patient deaths that occur within 7 days of a procedure. I look within 7 days of a procedure to minimize the possibility that a patient died for reasons unrelated to the surgeon, such as a nurse making a mistake during post-operative recovery. In Appendix Figure A5, I show that the results do not change if I restrict bad events to be deaths that occur on the same day of a procedure.

To identify good signals, I look at the top percentile of the riskiest patient-procedure

⁹See, for example, Tsugawa et al. (2016), Pine et al. (2007).

¹⁰ICD is the International Classification of Diseases coding.

pairs, where patient risk is defined as before and procedure risk is a procedure’s 30-day mortality rate, controlling for patient characteristics. The top 1% of pairs are those for which there is a high probability of death, either due to the patient’s risk, the procedure risk, or a combination. I call a patient outcome “unexpectedly good” if the patient does not die and is not readmitted to the hospital within 30 days of the procedure. While good events are rare by construction, 26% of surgeries that could be categorized as good end up being so. In the majority of cases, the patient is re-hospitalized. As a robustness check, I show how the results change with a 5% risk cutoff in Appendix Figure A6.

Propensity to Refer to Female Surgeons In the analysis, I test whether physicians with a high propensity to refer to women respond differently to signals than physicians with a low propensity to refer to women. To do so, I construct a measure of whether a physician over- or under-refers to women relative to the average physician in her HRR:

$$\pi_{j,s} = \frac{\text{Referrals to Women}_{j,s}}{\text{Total Referrals}_{j,s}} - \sum_{j' \in J} \frac{\text{Referrals to Women}_{j',s}}{\text{Total Referrals}_{j',s}} \quad (2)$$

Here, the first term is the number of physician j ’s referrals going to women in specialty s divided by physician j ’s total number of referrals to surgeons in specialty s . The second term is the total number of referrals going to women in specialty s from other physicians (j') in the same HRR divided by the total number of referrals from other physicians to surgeons in specialty s . If $\pi_{j,s} < 0$, the physician under-refers to women relative to other physicians in the same referral region and is classified as a “low propensity” physician. If $\pi_{j,s} > 0$, the physician over-refers to women relative to other physicians and is classified as a “high propensity” physician.

2.2.2 Summary Statistics

The list of surgical specialties considered in the paper and the gender distribution of each are presented in Figure 1. Women are under-represented, making up 17% of the full sample, but with heterogeneity by specialty.

Summary statistics for the unmatched sample of specialists who perform a surgery are presented in Table 1. In constructing this table, I reweight the sample so that men and women have the same specialty distribution. Panel A displays surgeon characteristics. All of the statistics are the average over one year. For example, the total number of patients is the number of patients that a surgeon sees on average in one year. Several important differences emerge. Female surgeons receive fewer total patients than male surgeons (81

compared to 112). Women's patients are less risky, younger, and more likely to be female or minority.

Patient deaths are relatively rare among the full set of surgeons. Approximately 0.85% of the patients that a male surgeon sees and 0.72% of patients a female surgeon sees in a given year die within 7 days of a procedure. Note that this does not account for differences in factors like patient risk. Good outcomes are, by construction, also quite rare when all referrals are taken as the base. In Table 1, I use the number of referrals a surgeon receives that qualify as risky procedures (i.e. the patients are in the top risk percentile) as the base. For both men and women, 26% of referrals they receive that qualify as risky will not result in a death or rehospitalization.

Mean surgeon ability is also reported. Women have a slightly higher mean ability and slightly lower variance of ability, but the differences are small. Figure 2 plots the probability distribution function and the cumulative distribution function of surgeon ability. The distributions for men and women largely line up with one another.

Panel B presents summary statistics for physicians, again measuring variables over the course of a year. Physicians refer to 15.5 different male surgeons and 2.5 different female surgeons per year. For any given procedure, physicians have relatively few surgeons whom they refer to, referring to 1.4 male surgeons per procedure and 0.2 female surgeons. Note that this is not the number of surgeons that are available to be referred to. Within the set of specialties that are included in the matched sample, physicians have an average of 7 available surgeons to refer to in a given specialty.

3 Empirical Strategy

To understand how physicians update their beliefs, I use an event study design to compare how referrals to male and female surgeons change after good and bad patient outcomes. The previous section showed that male and female surgeons receive different types of patients. I therefore use a matching procedure to match men and women who have similar characteristics and are performing the same surgery on similar patients. This section first describes the matching procedure and shows that the resulting sample is balanced in terms of both surgeon and patient characteristics. It then describes the event study design, estimating equations, and identification assumptions.

3.1 Matching Procedure

I carry out a coarsened-exact match for male and female surgeons on a set of patient and surgeon characteristics.¹¹ For this procedure, I start with the sample of all events, meaning that surgeons can be included in the sample multiple times if they experience multiple events. If a surgeon experiences multiple events with the same physician, I only keep the first instance of an event. Before matching, I have a sample of 265,734 distinct surgeon-physician pairs with bad events and 302,294 distinct pairs with good events. I then match male and female surgeons on variables measured over the four quarters before an event. I match with replacement of individual surgeons and physicians but allow each surgeon-physician pair to be matched only once.

I match exactly on the surgeon's specialty, the procedure being performed, and the patient's gender and minority status (white vs. non-white). Individuals for whom there is no exact match for each of these categories are dropped. Procedures are identified via the Current Procedural Terminology (CPT) code that the American Medical Association maintains to standardize medical service reporting. There are multiple "layers" of coding with 19 broad parent groups of surgical procedures. Within a parent group, there are three additional layers, each identifying a procedure more uniquely. In Appendix A.1, I give an example of what the layers look like for a sample parent group. I match on the second-finest layer, denoted by an asterisk in the appendix. Because there are slightly different procedures within each of these groups, I also match on the average risk of death for each procedure. I am therefore either matching on the exact procedure or on very similar procedures that differ slightly in treatment (for example, treating a humeral shaft fracture with plates/screws versus treating a humeral shaft fracture with rods).

I match coarsely on the patient's age and risk; the number and fraction of physician j 's referrals going to surgeon i prior to the event; the total number of referrals surgeon i received from any physician prior to the event; and surgeon experience, measured in terms of the number of years since the surgeon finished medical school and the number of specific procedures the surgeon had completed prior to the event. Finally, I match on a physician's outside option, using the number of other surgeons available in surgeon i 's speciality in the same HRR. I use an average of 9 bins for each variable.

Matching on patient age, gender, race, and risk attempts to minimize the differences between patients that the surgeons see. Matching on past surgeries, referrals, and years of experience allows me to compare two surgeons with similar experience levels, both

¹¹This procedure involves matching surgeons exactly on the surgeon's specialty, the procedure being performed, and the patient's gender and race. The other variables are then divided into bins and surgeons are matched across bins.

in performing the procedure at hand and in their overall experience. Matching on the number of surgeons in the same specialty as surgeon i in physician j 's referral region ensures that I am comparing physicians who have similar outside options. Note that I do not require the matched surgeons to have the same referring physician.

Tables 2 and 3 show that the final samples of matched male and female surgeons who experience bad and good events respectively are balanced. I end up with 7,757 surgeon matches for bad events and 6,979 matches for good events.

3.2 Estimating Equations and Identification Assumptions

I briefly overview the main estimating equations and identification assumptions that I use in the analysis. The remaining estimating equations are presented with the results in Section 4.

Estimating Equations and Identification for Matched Male/Female Surgeon Pairs

To identify the impact of a good or bad event on referrals, I designate the quarter in which an event occurs as quarter 0. I then sum the number of referrals a surgeon received from the referring physician in each quarter, starting four quarters before the event and ending six quarters after the event, and leaving out the patient referred for surgery. I stack the events of all physician-surgeon pairs and estimate equation 3 below, where $event_{ij,t-k}$ is a dummy variable indicating that an event occurred in quarter t and fem_i is a dummy variable indicating the surgeon is female:

$$R_{ijk} = \sum_{k=-4}^6 \beta_k event_{ij,t-k} + \sum_{k=-4}^6 \gamma_k (event_{ij,t-k} \times fem_i) + \theta_{ij} + \epsilon_{ijk} \quad (3)$$

The outcome variable, R_{ijk} , is the number of referrals physician j sends to surgeon i in quarter k . The coefficient $\hat{\gamma}_k$ tells us how a physician's reaction to an event changes when the surgeon is a woman.

I include physician-surgeon match fixed effects, θ_{ij} , to absorb any initial differences between matches and cluster standard errors at the physician-surgeon match level in case there are idiosyncratic factors that are specific to a particular physician-surgeon pair. This assumes that each pair's errors are uncorrelated with the errors of other pairs.

The main identification assumption is that women do not systematically receive different types of patients or perform different procedures than men. For example, if women receive less risky patients, it would make sense that a physician would update more about

them than about men after a bad event. Matching on patient characteristics, including patient risk, as well as on the procedure code should alleviate this concern. However, further robustness checks are presented in Section 5.

Estimating Equations and Identification for “Placebo” Surgeons

To understand what referral patterns between the physician and the surgeon would have looked like in the absence of an event, I create a set of placebo surgeons who perform a surgery but do not experience a good or bad event (e.g. they experience the expected outcome). To do so, I use the matching procedure described in Section 3.1 to match female surgeons who are identical on all observables and who receive observably similar patients, but one of whom has a patient die while the other does not. I do the same for male surgeons. For bad events, I am thus comparing surgeons who had a patient die (the treated surgeons) to those who did not (the placebo surgeons) and for good events I am comparing surgeons whose patient lived (treated) to those whose patient dies or is re-hospitalized (placebo). Balance tables are presented in Appendix Table C.

To quantify the impact an event has on referrals to a surgeon, I estimate

$$R_{ijk} = \sum_{k=-4}^6 \beta_k surgery_{ij,t-k} + \sum_{k=-4}^6 \gamma_k (surgery_{ij,t-k} \times event_k) + \theta_{ij} + \epsilon_{ijk} \quad (4)$$

on the matched female and matched male samples separately. The variable $surgery_{ij,t-k}$ is a dummy variable that equals one during the quarter of the surgery and $event_k$ is a dummy variable indicating that a good or bad event occurred. All other variables are defined as before.

Estimating Equations for Spillovers

In the analysis I test whether one surgeon’s performance affects how the physician updates about other surgeons. To do so, I estimate

$$f_{ijgsk} = \sum_{k=-4}^6 \beta_k event_{ij,t-k} + \sum_{k=-4}^6 \gamma_k (event_{ij,t-k} \times fem_i) + \delta_{available_{js}} + \theta_{ij} + \epsilon_{ijgsk} \quad (5)$$

separately on the samples of male and female surgeons. The outcome variable, f_{ijgsk} , is the fraction of physician j ’s new referrals going to surgeons in group g (men or women) in specialty s in quarter k , when surgeon i performs a surgery in quarter t . The measure leaves out referrals going to the performing surgeon i . Other variables driving the physician to refer less to women therefore have to change at the time of the event to explain the

changes in the physician’s behavior. For example, if other women in the physician’s HRR are less skilled or have full schedules, it is unlikely that the physician referred to them before the event occurred so the change in the fraction of referrals going to other women should not change.

I control for the fraction of available surgeons (*available_{js}*) who are of the same gender and in the same specialty as the performing surgeon to ensure that the results are not driven by constraints in physicians’ surgeon options. The result of this estimation tell us whether physicians change their beliefs about other female (male) surgeons after one of their patient lives or dies under the care of a female (male) surgeon.

Variables Correlated with Physician’s Reaction

To understand what drives a physician’s reaction, I correlate several variables with the change in referrals from a physician to a surgeon. Specifically, I estimate

$$R_{ijt} = \beta_1 Fem_i + \beta_2 Post_t + \beta_3 Var + \beta_4 (Fem_i \times Post_t) + \beta_5 (Fem_i \times Var) + \beta_6 (Post_t \times Var) + \beta_7 (Fem_i \times Post_t \times Var) + \sum_{X^k \in \mathbb{X}} X_{ijkt} + \theta_{ij} + \epsilon_{ijt} \quad (6)$$

where *Post_t* is a dummy equalling one for quarters after the event occurs and *Var* is the variable being tested for correlation with the physician’s response. A time trend and time trend interactions ($\sum_{X^k \in \mathbb{X}} X_{ijkt}$) with the variables $\mathbb{X} = \{Post_t, Fem_i, Var\}$ are also included.

4 Results

The empirical analysis uncovers two asymmetries in how physicians react to signals. First, there is an individual-level asymmetry. Physicians respond more to good signals if the surgeon is male and more to bad signals if the surgeon is female. Comparing surgeons who experienced a good or bad outcome with the placebo surgeons who did not, I find that while the referral path for men and women would have been similar had the event not occurred, women receive far fewer referrals following a bad outcome than they otherwise would have. Conversely, men receive more referrals after a good outcome than they otherwise would have. Second, there is an asymmetry at the group level. Physicians update their beliefs about other women after receiving a signal from an individual woman but do not update about other men after receiving a signal from one man.

Physicians who just began a referral relationship with a particular surgeon react the

strongest. The longer is a physician’s referral relationship with a surgeon, the less that physician reacts to patient outcomes. Further, the gender gap in physicians’ responses is decreasing in the length of the referral relationship. In what follows, I therefore focus primarily on physicians and surgeons who just began a referral relationship (i.e. where the physician had sent fewer than 10 referrals prior to the event), as the asymmetric response is concentrated among these physicians and surgeons. However, I show how the results depend on the length of the referral relationship in Section 4.2.1.

4.1 Responses to Signals

4.1.1 Updating after a Bad Event

This section documents the referring physician’s reaction to a patient death. Because I match male and female surgeons on both the number and fraction of referrals they received from a physician before an event occurs, I am comparing two surgeons for whom physicians have roughly the same beliefs.¹²

Figure 3 documents how physicians respond to a patient death, plotting the quarterly coefficients $\hat{\beta}_k$ and $\hat{\beta}_k + \hat{\gamma}_k$ for $k \in \{-4, 6\}$ from estimating equation 3. In this figure, physician j refers patient p to surgeon i . Surgeon i performs the surgery in quarter 0 and the patient dies within 7 days of the surgery. Coefficients are plotted relative to the number of referrals the surgeon received from physician j in the quarter before the event ($k = -1$). Both male and female surgeons were receiving 0.65 referrals in $k = -1$ from their referring physician.

The figure shows that before the patient death, the surgeon receives an increasing number of referrals from the physician as they have just started their referrals relationship. When the death occurs in quarter 0, the physician refers less to the surgeon, with a marked drop in referrals occurring if the surgeon is female. Controlling for time trends, men receive 0.101 more referrals in the quarters after the death relative to the quarter before the death, whereas women receive 0.222 *fewer* referrals after the death relative to the quarter before.¹³ The gap in referrals that emerges between male and female surgeons is significant at the 1% level and persists up to a year and a half following the death. To put these numbers into perspective, a male surgeon would have to have three patient deaths to be treated the same as a female surgeon. Given the rarity of patient deaths, this difference is substantial. The results are summarized in column 1 of Table 4.

¹²This is assuming that referrals are a function of a physician’s belief about a surgeon’s average ability, a claim that I further explore in Section 7

¹³See Table 4, column 1.

I also find a change in the types of patients and procedures that the physician refers in the future. In Figure 5, I re-estimate equation 3 but use the average risk of patients that physician j refers to surgeon i in quarter k as the dependent variable. There is no significant change in the riskiness of patients that physicians send to men. However, women start to receive patients that are 0.28 standard deviations less risky after a death. Similarly, women are referred procedures that are 0.041 standard deviations less risky than what they received before the death. The impact that this has on surgeon pay and skill accumulation is further discussed in Section 6

To further quantify the impact that a death has on male and female surgeons, I show what the referral path would have looked like in the absence of a death using the matched sample of treated/placebo surgeons described in Section 3. The matched surgeons are identical on observables and receive patients with similar risk levels and demographics, but one surgeon experiences a patient death while the other does not. It is worth noting that the absence of a death is not equivalent to the physician receiving no signal from a surgeon. The fact that the patient lives is in itself a good signal. I am therefore not comparing a surgeon who sends a bad signal to one who does not send a signal.

Figure 4 plots the results from estimating equation 4 separately for male and female surgeons. In both figures, the estimates for the placebo surgeons indicate that the number of referrals to male and female surgeons would have continued to grow at a similar rate had the patient not died. However, the gap between the actual and “projected” number of referrals is smaller for men (top panel). Men who experience a patient death receive 0.22 fewer referrals each quarter than they would have if the patient lived. The F-statistic for the joint significance of $\hat{\delta}_1 - \hat{\delta}_6$ is 2.48. Women who experience a patient death, shown in the bottom panel, receive 0.6 fewer referrals each quarter (with an F-statistic of 15.88). Physicians thus seem to update their beliefs downward about both male and female surgeons, but by a larger amount when the surgeon is female.

The difference in outcomes between women and men who do and do not experience a patient death is summarized in a difference-in-differences plot shown in panel (c) of Figure 4. There is no difference in the referral path between male and female surgeons who do not experience a patient death (grey circles) but women who experience a patient death receive approximately 0.3 fewer referrals per quarter than men who experience a death (green triangles).

4.1.2 Updating after a Good Event

It is possible that physicians react more to bad signals from women because the variance of women’s ability is larger. If this is true, physicians should react strongly to good signals

from women as well.¹⁴ I now test how physicians react to “unexpectedly good” outcomes, defined as the top 1% of the riskiest patient-procedures pairs for which no death or re-hospitalization occurred. The quarterly coefficients from estimating equation 3 as well as the placebo outcomes are presented in Figure 6.

Opposite to their reaction to a death, physicians respond more strongly when the surgeon is male than when the surgeon is female. While referrals to both male and female surgeons increase, they increase by a significantly larger amount for male surgeons. Controlling for time trends, men receive roughly 0.6 more referrals than they did in the period before the good event whereas women receive 0.35 more referrals, a difference that is significant at the 5% level (see Table 4, column 2 for a summary of effects). A physician’s response to good signals stands in stark contrast to how physicians react to bad signals, where their response to male surgeons is muted.

Relative to the sample of surgeons who did not experience a good outcome, men receive 0.25 more referrals per quarter than they otherwise would have while women receive 0.15 more. Note that a difference in the number of referrals to men and women who do not experience a good outcome also emerges as the expected outcome in this case is a hospital readmission or death. Although a re-hospitalization is expected, women still receive slightly fewer referrals than men do after the surgery.¹⁵ The difference-in-differences coefficients are plotted in the bottom panel of Figure 6.

4.1.3 Spillovers: Updating about Other Surgeons

I now turn to the question of whether a physician’s experience with one surgeon influences that physician’s beliefs about other surgeons of the same gender. Figure 7 plots the quarterly coefficients from estimating equation 5 where the event is a patient death.

The outcome in the top figure is the fraction of a physician’s referrals going to surgeons the physician hasn’t referred to before who are the same gender and in the same specialty as the performing surgeon. I focus on new referral relationships as a physician’s experience with one surgeon does not significantly impact her beliefs about another surgeon she has been referring to for some time. This is shown in Appendix Figure A1 where physicians who have a long referral history with a particular surgeon do not change their referrals to that surgeon after having a bad experience with surgeon i .

In cases where a female surgeon has a patient die (the red triangles), the physician be-

¹⁴This is assuming that ability is distributed according to a symmetric distribution. Such assumptions are formally analyzed in Section 7

¹⁵In Section 4.2.3, I show how physicians’ responses to deaths depend on patient risk as well as several other factors.

comes less likely to form referral relationships with female surgeons in the future. However, physicians do not change their propensity to form new referral relationships with male surgeons after a patient dies under a man (the blue circles). Men appear to be treated as individuals while information about one woman in specialty s affects the physician's beliefs about other women in specialty s .

The bottom figure shows the change in a physician's referrals to new surgeons of the same gender but in different specialties than the performing surgeon. Physicians slightly reduce the fraction of their referrals going to female surgeons in other specialties but the post-death coefficients are jointly insignificant.

The spillover results are summarized in columns 3 and 4 of Table 5. Relative to the mean fraction of referrals going to new female surgeons in the same specialty, the decline in referrals is substantial. The fraction of a physician's referrals going to new women in the same specialty as the performing surgeon declines by 53% (column 1). The fraction going to new women in other specialties (column 3) declines by 20% (but again, this result is insignificant).

Interestingly, physicians do not appear to treat women as a group when a woman performs well. Column 2 of Table 5 shows the spillovers to other women after a female surgeon has a good patient outcome. The coefficient on the $\text{female} \times \text{post}$ interaction variable is positive but insignificant.¹⁶ It is possible, though, that a physician updates her beliefs about other female surgeons upward but continues to only refer to the performing surgeon.

These results also provide evidence that the drop in referrals is not simply due to female surgeons changing their behavior. A large literature shows that women are less likely to be overconfident than men (Lichtenstein, Fischhoff, and Phillips, 1982; Beyer, 1990; Barber and Odean, 2001). It is possible that female surgeons turn away more referrals after a patient death if the event hurts their confidence. The fact that physicians are changing their behavior toward other female surgeons, though, suggests that at least part of the change is due to physicians.

4.1.4 Information Spillovers to Other Physicians

The implications that asymmetric updating has on a surgeon's career depends in part on whether information about an event spreads to other physicians. I test whether other

¹⁶Appendix Figure A2 plots the coefficients from estimating equation 5 using good events. There are no significant positive spillovers to other women or men in the same specialty as the performing surgeon.

physicians react to a bad event in Figure 8. I plot the coefficients from estimating

$$R_{i,-j,k} = \sum_{k=-4}^6 \beta_k event_{ij,t-k} + \sum_{k=-4}^6 \gamma_k (event_{ij',t-k} \times fem_i) + \theta_{ij} + \epsilon_{ijk} \quad (7)$$

where $event_{ij,t-k}$ is still a dummy variable indicating that a bad event occurred between physician j and surgeon i , but the outcome variable is the number of referrals that surgeon i receives from other physicians (excluding j). I consider two outcome variables: the number of referrals going from other members of physician j 's group practice and the number of referrals from physicians outside of physician j 's practice.

Panel (a) plots the coefficients when the outcome variable is the number of referrals to the performing surgeon from other members of physician j 's group. I restrict the group practices to be those containing at most 20 members. This is done to account for the fact that many group practices defined in the Medicare data cover groups with branches in multiple regions. It is therefore unclear whether group practices with many members are large practices in one geographic location or normal-sized practices with multiple branches. There is a small decline in the number of referrals from other members of small group practices. Female physicians receive on average 0.5 fewer referrals per quarter after a bad event than before, but the coefficients in quarters 1 through 6 are jointly insignificantly different from zero. Male surgeons, on the other hand, continue to receive referrals from the referring physician's practice.

Panel (b) shows the change in referrals from physicians outside of physician j 's practice but in the same HRR. Here we see no impact on the number of referrals the surgeon receives. Both women and men continue to receive referrals from physicians outside of the referring physician's practice, suggesting that information does not spread or that physicians do not act on this information.

4.2 What Influences a Physician's Reaction to a Signal?

I have shown that physicians respond differently to a given patient outcome depending on the surgeon's gender. Physicians lower referrals to female surgeons more than male surgeons after a bad outcome and increase referrals more to male surgeons than to female surgeons after a good outcome. I now explore other factors that influence a physician's reaction, finding two main drivers. First, physicians' reactions to signals are weaker the more signals they have received from a surgeon in the past. Physicians who have received many signals from a woman prior to the event are also more likely to treat her the same as a man. Second, physicians with a high propensity to refer to women before the event treat

men and women more equally - e.g. exhibit less asymmetric updating - than physicians with a low propensity to refer to women.

4.2.1 Referral History with Surgeon

So far I have restricted the analysis to physician-surgeon pairs in which the physician has referred at most 10 patients to the surgeon before the event. This concentrates the analysis on physicians who are presumably learning about a particular surgeon. If physicians learn about surgeons over time, they will become more certain in their beliefs about a surgeon's ability the longer they have been referring to one. To test how a physician's reaction changes with the length of her relationship with a surgeon, I estimate equation 6 where the independent variable of interest is the number of referrals that physician j sent to surgeon i in the year before the event.¹⁷ I estimate equation 6 on the set of matched surgeons who have had up to 50 prior referrals from a given physician. I am still comparing male and female surgeons who receive the same number and fraction of referrals before an event occurs, but am looking at how a physician's reaction changes the more a physician has worked with a surgeon.

The results are presented in columns 1 and 3 of Table 6. *Pre-Referrals* is the number of patient referrals that physician j sent to surgeon i before the bad (column 1) or good (column 3) event. These "pre-referrals" are referrals that ended well: the patient did not die and was not readmitted to the hospital. I am therefore measuring how an additional referral that ended in a good patient outcome affects the physician's response. An additional referral from a physician to a surgeon reduces the physician's negative response to a death by 0.014 referrals per quarter for male surgeons and 0.022 referrals per quarter for female surgeons, thereby diminishing the gender gap in physician response. Female surgeons need to have sent 22 good signals (performed 22 procedures that went well) to a physician in the year before the event for there to be no gender gap in the physician's response.

An additional referral mitigates a physician's positive reaction to a good outcome when the surgeon is male but does not change the positive response when the surgeon is female, again leading to more equal responses. These results suggest that physicians learn about surgeons, becoming more certain in their beliefs about a surgeon's ability over time. Any bias that they exhibit comes out only when working with new surgeons.

¹⁷I do not look at the full referral history as the data is cut off in 2008. However, these two measures should be correlated.

4.2.2 Physician Response and Referral History with Other Surgeons

The previous section showed that the longer the referral relationship between a physician and a surgeon, the more muted the physician's response to an event is. It is unclear, though, how a physician's response is influenced by her referral history with other surgeons of the same gender as the performing surgeon. For example, a physician who has favorable beliefs about women and who sends a large volume of referrals to them presumably has more information about women. That physician may not react as much to new signals. Depending on the shape of the physician's priors, though, a surgeon who thought women were very good and was consequently surprised to see a patient die under a woman might drastically revise her beliefs.

To understand how a physician's response varies with her referral history with other surgeons, I again estimate equation 6 and use the "propensity to refer to women" variable described in Section 2 as the main explanatory variable. A physician's propensity to refer to women is measured as the difference between the fraction of her referrals going to women in a particular specialty and the average fraction going to women of other physicians in the same referral area. I then define a variable, *High Propensity*, that equals one if a physician sends a greater fraction of her referrals to women within a given specialty than the average physician in her referral region.

The results are presented in columns 2 and 4 of Table 6. Physicians with a high propensity to refer to women react less negatively after bad outcomes and more positively after good outcomes. The gender gap in the physician's reaction shrinks by 22% if a physician has a high propensity to refer to women.

4.2.3 Other Variables Influencing Physician's Reaction

I also test whether a physician's outside option, surgeon and physician experience, and physician gender influence the physician's reaction to a signal.

Outside Options It is plausible that a physician who has many possible surgeons she can refer to would have a stronger response to a patient death as it is easier for her to shift to a new surgeon. To test whether a physician's outside option matters, I use the number of surgeons in the same specialty as the performing surgeon in the referring physician's HRR. The results from estimating equation 6 are presented in Column 1 of Table 7, the x-variable of interest is physician j 's outside option. There is a small impact of having more outside options but the result is statistically insignificant.

Experience I look at both the physician’s and the surgeon’s experience, measured as the number of years since they graduated from medical school, to understand whether experience influences a physician’s reaction to signal. I again estimate equation 6 but include $Exper_{i,t}$ (the surgeon’s experience at the time of the event) or $Exper_{j,t}$ (the physician’s experience) as the main independent variable. I also include squared experience terms to allow for non-linearities in the doctor’s response. The results are presented in columns 2 and 3 of Table 7. There is no significant relationship between either doctor’s experience and the physician’s reaction. A physician’s experience working with a particular surgeon seems to matter more than absolute years of experience.

Physician Gender The evidence on how women evaluate other women is mixed. For example, Casadevall and Handelsman (2014) find that having a woman on a convening team for scientific conferences increases the proportion of invited women. Bagues et al. (2017), however, find that women evaluating tenure cases do not favor female candidate and that men in fact become harsher in their evaluations in the presence of a female evaluator.

I find no significant evidence that male and female physicians treat female surgeons differently. In column 4 of Table 7, a female dummy indicating that the physician is female is included as the independent variable of interest. The point estimate on the triple interaction between *Female*, *Post*, and *Female physician* is positive, suggesting that female physicians might be easier on female surgeons than male physicians, but the result is noisy and I cannot rule out no or negative effects.

Summary of Results

In sum, I look at the impact of two signals, good and bad, on two sets of outcomes for male and female surgeons: referrals to the performing surgeon and referrals to other surgeons of the same gender. The following table, which draws from Tables 4 and 5, shows the average post effect of an event on referrals to individual surgeons and to other surgeons.

	Performing Surgeon		Other Surgeons	
	Male	Female	Male	Female
Bad Outcome	0.101	-0.222	∅	-0.079
Good Outcome	0.604	0.346	∅	∅

5 Alternative Interpretations

Before discussing what the empirical findings tell us about how physicians update their beliefs, I explore three alternative interpretations of the results. I show that differences in unobservable risk, differences in the predictiveness of events for future events, and changes in the physician's behavior cannot account for the findings.

5.1 Differences in Patient Risk

Although I match on observable patient risk, there could be unobservable factors influencing risk. If female surgeons receive less risky patients, deaths are surprising and survivals are unsurprising, meaning that physicians respond strongly to deaths and weakly to survivals. Here, I put bounds on what the unobservable risk difference between male and female surgeons' patients would have to be to justify the degree of differential updating.

In Figure 9, I plot the regression coefficients $\hat{\beta}_1$ and $\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$ from estimating

$$\Delta R_{ijp} = \beta_1 Risk_p + \beta_2 (Risk_p \times Fem_i) + \beta_3 Fem_i + \theta_i + \epsilon_{ijp} \quad (8)$$

where ΔR_{ijp} is the change in referrals from physician j to surgeon i before and after patient p 's death, and $Risk_p$ is the patient p 's observed risk level. The coefficient $\hat{\beta}_2$ tells us what the difference in unobserved risk between a female surgeon's and a male surgeon's patient would have to be to account for the gender difference in the physician's reaction.

While physicians respond to patient risk regardless of surgeon gender, they respond much more when the surgeon is female. Figure 9 shows that a male surgeon with a patient in the bottom risk decile experiences a drop in referrals equivalent to what a female surgeon with a patient in the 7th decile receives. Thus, any differences in unobserved risk would have to be large enough to move a patient that a male surgeon sees from being in the bottom 10th percentile of patient risk to the top 70th percentile of observed risk, a large difference.

5.2 Are Outcomes Differentially Predictive of Future Outcomes?

A physician's behavior is also be justified if bad events are predictive of future bad events for female surgeons and good events are predictive of future good events for male sur-

geons. To test this hypothesis, I estimate

$$\mathbb{P}(Event_i = 1|X_i, X_{pt}, X_{i,p'}) = \beta_1 Fem_i + \beta_2 FutRefs_i + X_i' \gamma + \alpha \log(PastRisk_p) + \delta \log(FutRisk_{p'}) + \epsilon_{ip}$$

on the matched samples of surgeons who experienced a bad or good event. The outcome variable is the probability that surgeon i has another event in the future under any physician (not just physician j). I condition on the number of referrals the surgeon receives in the future from all physicians ($FutRefs_i$) as well as the log of future patient risk ($\log(FutRisk_{p'})$). I also control for the performing surgeon's characteristics (X_i'), including experience, specialty, and work history¹⁸, as well as the log of the surgeon's past patients' risk levels ($\log(PastRisk_p)$).

The results are presented in Table 8. I find no evidence that patient deaths are more predictive of future deaths for female surgeons than for male surgeons. In fact, women are less likely to have future patients die even conditional on the risk of future patients. If physicians are responding to an unobservable factor, it is not something that influences the future performance of surgeons.

5.3 Do Physicians Stop Referring for Certain Procedures?

An additional concern is that physicians stop referring patients for a particular surgery after a patient dies. For example, Keating et al. (2017) find that doctors who refer a patient for a colonoscopy are less likely to refer patients for that procedure if the patient has an extreme adverse outcome. However, the effect is short-lived as physicians begin referring for the procedure again, and at the same rate as before, one quarter after the adverse outcome. Nevertheless, if women tend to specialize in one type of surgery while men perform many different surgeries, the drop in referrals to women could be due to the physician changing the types of procedures she refers.

I test this by looking at how referrals for the surgery that was performed change after a patient death. Specifically, I estimate

$$S_{jk} = \sum_{k=-4}^6 \beta_k event_{ij,t-k} + \sum_{k=-4}^6 \gamma_k (event_{ij,t-k} \times Fem_i) + \theta_{ij} + \epsilon_{ij} \quad (9)$$

where S is the surgery that was being performed on the patient who died and S_{jk} is the number of those surgeries that physician j refers to any surgeon in quarter k .

¹⁸This includes the number of patients seen before the death.

The results are shown in Appendix Figure A3. For female surgeons, the coefficients on all quarters before and after the event are precise zeros, indicating that the physician does not change her referral patterns for the surgery in question. For male surgeons, there is a small drop in the number of referrals the physician gives for the procedure in question in $k = 1$ but the physician’s referrals revert back to the mean shortly thereafter.

I also show that the physician decreases referrals to surgeon i for all procedures, not just the one that was being performed on the patient who died. In Appendix Figure A4, I plot the coefficients from estimating

$$OtherRef_{ijk} = \sum_{k=-4}^6 \beta_k event_{ij,t-k} + \sum_{k=-4}^6 \gamma_k (event_{ij,t-k} \times Fem_i) + \theta_{ij} + \epsilon_{ij} \quad (10)$$

where the outcome variable, $OtherRef_{ijk}$, is the number of referrals that the physician sends to the surgeon aside from the procedure that was being performed on the patient who died. The results are noisy as physicians typically only refer to a surgeon for a particular procedure, but the patterns are the same. Referrals for other procedures drop for both men and women but by a greater amount for women. The coefficients for female surgeons are significantly different from those for male surgeons at the 10% level.

6 Welfare Analysis and Career Effects

6.1 Surgeon Ability

If asymmetric updating distorts a physician’s belief about a surgeon’s ability, physicians may switch away from high ability female surgeons after receiving negative signals. For example, if physicians have some cutoff ability, below which they do not refer to a surgeon, physicians will stop referring to female surgeons earlier than similar men. The average ability of male surgeons they refer to will eventually be lower than that of the female surgeons they refer to.

To test whether asymmetric updating affects the average quality of surgeons a physician refers to, I use the definition of surgeon ability described in Section 2. I then calculate the average ability of all surgeons a physician refers to in each quarter and plot how it changes after a bad event under one surgeon. If physicians give male surgeons too many chances to make mistakes and female surgeons too few chances, average surgeon ability will decline after a patient death as physicians move away from potentially qualified female surgeons to potentially less qualified male surgeons.

Figure 10 plots the quarterly coefficients from estimating

$$\bar{a}_{jk} = \sum_{k=-4}^6 \beta_k event_{ij,t-k} + \sum_{k=-4}^6 \beta_k (event_{ij,t-k} \times fem_i) + \theta_{ij} + \epsilon_{ijk}$$

where \bar{a}_{jk} is the average ability of the surgeons that physician j refers to in quarter k . There is no significant change in the quality of surgeons that physicians refer to when the performing surgeon is a man. However, when the performing surgeon is a woman, the average surgeon quality falls by approximately 0.1 standard deviations in the year following a death, although these results are only marginally significant at the 10% level. However, they provide suggestive evidence that physicians may be misestimating female surgeons' abilities after deaths and switch from some high ability female surgeons to other lower ability surgeons.

6.2 Surgeon Pay and Skill Accumulation

I now turn to the impact that asymmetric updating has on surgeons' career trajectories. In Section 4, I find limited evidence that information about a patient death spreads to other physicians. There is a small decline in the number of referrals that women receive from other members of the referring physician's group practice, provided the practice is small. However, this decline is statistically indistinguishable from zero. As such, I abstract from changes in other physicians' behavior and focus on the impact that asymmetric updating by the referring physician has on pay and skill accumulation.

Numerous papers have documented a pay gap between male and female physicians.¹⁹ Closely related to this paper, Zeltzer (2017) decomposes the Medicare earnings gap, showing that in the raw data, women earn 48% less than men. About a third of the gap can then be explained by difference in the specialties men and women select into. Controlling for career interruptions and differences in experience and education, Zeltzer shows that gender homophily in referrals explains an additional 15% of the gap. However, the remainder of the gap remains unaccounted for.

Here, I provide an additional mechanism that contributes to the gap, but do not quantify the contribution. It is difficult to estimate the full impact that asymmetric updating has on the surgeon pay gap as I look at two specific and relatively infrequent events. Surgeons experience less than one bad outcome per year, for example. Furthermore, physicians exhibit asymmetric updating when they are starting new referral relationships, so the relevant baseline pay gap for comparison is the gap that exists between men and women at

¹⁹See, for example, Ly et al. (2016), Lo Sasso et al. (2011), and Sasser (2005)

the beginning of their careers or who have moved. I therefore show that asymmetric updating creates a wedge between male and female pay but do not speculate on its relative importance in explaining the overall surgeon pay gap.

Figure 11 shows the change in Medicare payments following a patient death, plotting the quarterly coefficients from estimating

$$Pay_{ijk} = \sum_{k=-4}^6 \beta_k Death_{ij,t-k} + \sum_{k=-4}^6 \gamma_k (Death_{ij,t-k} \times fem_i) + \theta_{ij} + \epsilon_{ijk} \quad (11)$$

on the sample of matched surgeons who experience a patient death. The outcome variable, Pay_{ijk} , is the total quarterly Medicare pay that surgeon i receives from physician j 's referrals in quarter k . A patient death occurs in $k = 0$.

Because I have matched on a number of variables that influence the gender pay gap (such as volume of referrals, experience, and specialty), there is a small but statistically insignificant difference in male and female surgeon pay before the death. After the death, a gap of approximately \$140 per quarter emerges. Column 3 of Table 4 summarizes the effect, showing that women lose approximately 60% of their Medicare billings from the referring physician while men lose 30%. Women incur a substantial pay penalty from the referring physician but given that other physicians do not change their behavior much, women do not experience a substantial pay penalty overall.

Column 4 of Table 4 shows the impact of a good patient outcome on Medicare pay. Men receive a 36% increase in quarterly Medicare billings while women receive a 19% increase, although the difference is statistically insignificant.

A second channel through which asymmetric updating can influence women's careers is through skill acquisition. In Section 4, I showed that female surgeons who still receive referrals after a patient death receive easier cases, either with less risky patients or less risky procedures. Since learning-by-doing is important for surgeon learning²⁰, asymmetric updating might impact women's skill accumulation which can also influence future pay as well as their career trajectory.

7 Theoretical Framework

This section sets up a theoretical framework to link the empirical results to belief updating, answering whether and under what conditions the observed behavior is in line with Bayesian updating.

²⁰See, for example, Hughes (1991), Keehner et al. (2006).

To be consistent with the main results, a model must have two key features:

1. *Asymmetry in Updating about the Individual*: Physicians must update their beliefs more about men after a good signal and more about women after a bad signal.
2. *Asymmetry in Updating about Groups*: Physicians update their beliefs about other women upon receiving a signal from one woman but do not update beliefs about other men after receiving a signal from one man.

I do not argue that one particular model explains the results, but rather outline the assumptions about a physician’s beliefs that are needed for the behavior to be consistent with Bayesian updating. To derive these assumptions, I model the physician’s decision problem to map referrals to beliefs. I then show that the behavior is consistent with Bayesian updating if (1) physicians believe that women are higher ability than men, or (2) the difference in the average variance of women’s and men’s abilities increases as physicians receive more signals. I discuss the empirical validity of these assumptions and show that if physicians hold such beliefs, they are inconsistent with the data on the distribution of surgeon ability, allowing me to reject rational expectations. I then discuss an alternative model in which physicians exhibit bias dynamically rather than through their priors.²¹

7.1 Physician’s Decision Problem

Setup

I follow the setup in Zeltzer (2017). There are two types of agents, physicians and surgeons. Physicians, denoted $j \in J$, decide which surgeon, $i \in I_j$, to refer a patient to where I_j is the pool of surgeons available to j . Surgeons belong to an identifiable group $g \in \{m, w\}$ (men or women) and their ability, a_i , is unknown to the physician. In period $t = 0$, physicians have a prior probability distribution over a surgeon’s ability $a_{i,t} \sim f(\bar{a}_i, \sigma_i^2)$ where \bar{a}_i is the physician’s prior about surgeon i ’s average ability. In time 0, the physician’s prior is based on her beliefs about the group: $\bar{a}_{i,0} = \bar{a}_g$. Similarly, σ_i^2 is the variance and $\sigma_i^2 = \sigma_g^2$ in $t = 0$. I assume that f has a defined mean and variance but do not place restrictions on higher-level moments.

After receiving a patient, surgeons draw and send a signal (a patient outcome). To match the data, I assume that signals can be either good or bad: $s \in \{s_G, s_B\}$.²² The

²¹Under some assumptions, various other models also fit the data. See, for example, Bordalo et al.’s (2016) model of stereotypes.

²²The model can be extended to include a set of finitely many ordered signals and the results do not change.

probability of drawing each type of signal depends on a surgeon's ability, with higher ability surgeons being more likely to draw a good signal. Specifically, let the probability that a member from group g draws a signal s be $\mathbb{P}(s|g) = \sum_{i \in g} \mathbb{P}(s|a_i) f(a_i) da$.

Physicians want to maximize patient utility by referring to the best available surgeon subject to idiosyncratic factors like patient preferences and wait times. That is, physician j chooses a surgeon i to maximize

$$\operatorname{argmax}_{i \in I_j} U_p(a) = \beta \mathbb{E}[a_{i,t} | \bar{a}_i, g, s] - \lambda \sigma_{i,t}^2 + \epsilon_{ipt} \quad (12)$$

where \bar{a}_i is the physician's prior about surgeon i 's mean ability, and where i belongs to group $g \in \{m, w\}$. The constant λ represents risk preferences and ϵ_{ipt} represents the idiosyncratic factors discussed above. The physician is thus trying to choose the surgeon with the highest expected ability, the first term, while trading off the variance of ability, the second term.

Assuming ϵ_{ipt} is independently and identically distributed according to the extreme value distribution, we can write equation 12 in terms of the logit probability,

$$\mathbb{P}(R_{i,p,t}^j = 1 | a_{i,g,t-1}) = \frac{e^{\nu_{ij}}}{\sum_{i' \in I} e^{\nu_{i'j}}} \quad (13)$$

where $R_{i,p,t}^j = 1$ if physician j refers patient p to surgeon i in time t , and $\nu_{ij} = \beta \mathbb{E}[a_{i,t} | \bar{a}_i, s] - \lambda \sigma_{i,t}^2$. Physician j 's total number of referrals to surgeon i in period t can then be written as

$$R_{ijt}^{\text{total}} = n_t \cdot \frac{e^{\nu_{ij}}}{\sum_{i' \in I} e^{\nu_{i'j}}}$$

where n_t is the number of patients that the physician refers to surgeon i in time t if the surgeon's capacity constraint has not been reached.

Aside from idiosyncratic factors like patient preferences, two main variables influence the physician's referral choice in this model: beliefs about ability ($\mathbb{E}[a_{i,t}]$) and the variance of ability ($\sigma_{i,t}^2$). Referrals are increasing in the surgeon's expected ability and decreasing in the surgeon's expected variance of ability.²³

I consider each case in turn. I first look at how Bayesian physicians react to signals if they only care about mean ability, placing no restrictions on higher-order moments of the ability distribution. I then consider how physicians react if they have risk preferences and

²³This assumes risk aversion in which a physician prefers to refer to a surgeon who produces average outcomes for sure than one who produces either very good or very bad outcomes with some probability. There may be cases in which a physician would prefer to refer to a surgeon with a high expected σ^2 instead of a "safe" option, but I abstract from this case here.

care about the variance of ability.

7.2 Bayesian Updating: Physician Cares about Mean Ability

Recall that physicians have the prior probability distribution $a_{i,t} \sim f(\bar{a}_i, \sigma_i^2)$ over a surgeon's ability. If a physician only cares about mean ability, she chooses i to maximize $U_p(a) = \beta \mathbb{E}[a_{i,t} | \bar{a}_i, g, s] + \epsilon_{ip}$, which is equation 12 setting $\lambda = 0$. Holding the physician's beliefs about other surgeons constant, the change in referrals after a signal is then

$$\Delta \bar{R}_{ijt} = \frac{(n_t - n_{t-1}) \cdot e^{\beta(\mathbb{E}[a_{i,t} | \bar{a}_i, g, s] - \bar{a}_i)}}{\sum_{i' \in S} e^{\beta a_{i', t-1}}} \quad (14)$$

Empirically, the change in referrals to women is larger after a bad signal and smaller after a good signal compared to the change in referrals to men. For this to be true, it must be that $\mathbb{E}[a_{i,t} | s, w] - \bar{a}_{i,w} \leq \mathbb{E}[a_{i,t} | s, m] - \bar{a}_{i,m}$ in equation 14. This is possible under Bayesian updating when physicians believe that women have higher average ability than men.

Consider the following proposition:

Proposition 1. *If physicians are Bayesian and the following statements are true:*

1. $\mathbb{E}[a_i | s_G, g] > \mathbb{E}[a_i | s_B, g]$
2. $\mathbb{E}[a_{i,t} | s, w] - \bar{a}_{i,w} \leq \mathbb{E}[a_{i,t} | s, m] - \bar{a}_{i,m}$

then it must be that $\frac{\mathbb{P}(s_G | w)}{\mathbb{P}(s_G | m)} > 1$.

Proof. See Appendix B. □

Proposition 1 states that if expected ability is increasing in the signal and that the change in beliefs about women's ability are always more negative than the change in beliefs about men's ability, then women must have a higher average ability than men, which is equivalent to women having to send more good signals. Intuitively, if a Bayesian physician believes that women are more likely to send good signals, any good signal is going to be close to the physician's prior and will not move her beliefs much. In contrast, when a man sends a good signal, the signal is far from the physician's prior, leading her to update more. Similarly, if a woman sends a bad signal, the physician will update much more than she does about a man, whom the physician already believes to be low ability.

Discussion of Assumptions

If physicians believe that women are higher ability than men and physicians only care about mean ability, equation 12 implies that women should receive at least as many referrals as men do. Empirically, female surgeons receive 10% of referrals while they make up 17% of the surgeon population. Adjusting for the fact that women work fewer hours than men, they should receive at least 15% of referrals if physicians refer to men and woman at equal rates.²⁴ While the difference is small, female surgeons are under-referred to even after adjusting for work hours.

Theoretically, a cost of referring to women could result in men and women receiving an equal number of referrals but women being higher ability than men. A female surgeon would have to be sufficiently high ability to offset any cost associated with referring to her. Such costs could arise for a variety of reasons. For example, in a model of taste-based discrimination, physicians pay a utility cost to refer to women so they will only refer to women whose ability outweighs this cost.²⁵

Two empirical observations suggest that physicians do not believe that women are higher ability than men. First, in Table 6, I showed that physicians who send a larger fraction of their referrals to women react less to bad events and more to good events. If physicians who refer more to women than average have higher beliefs about women's ability, we should see the opposite result. Beliefs about women's ability should be negative correlated with referrals after a bad event as these are the physicians who should be the most surprised to see a bad outcome. However, these are empirically the physicians who react strongly to good signals and weakly to bad signals.

Second, while I cannot observe beliefs directly, I can look at the ability distributions of men and women to test for differences in average ability and to test whether physicians have rational expectations. Women are slightly higher ability than men. Figure 2 shows the ability PDF and CDF for male and female surgeons in the unmatched sample. Table 1 reports the means. Women have higher average ability but the difference is small: women's ability is 0.0001 points larger than men's. It is unlikely that such a small difference in ability leads to such a large reaction from physicians. Nevertheless, because I do not directly observe physicians' beliefs, it is possible that they believe that there is a large difference in surgeon ability, which produces the observed asymmetries in referrals. If this is the case, physicians' actions are in line with Bayesian updating, but are inconsistent

²⁴See Appendix A.1 for description and calculation of work hours adjustment

²⁵Similarly, women might work less convenient hours or be geographically distant from patients, making them more costly to refer to. Physicians would only refer to surgeons if their ability was sufficiently high to offset these costs

with physicians having rational expectations as their beliefs would not match the actual surgeon ability distribution.

7.3 Bayesian Updating: Physician Cares about Variance

If physicians also care about the variance of ability, they might choose to refer to a surgeon with a lower average ability but also a lower variance as this would mean less uncertainty in the patient's outcome. For example, if physicians are risk averse and want to avoid very bad outcomes like deaths, they might refer to a surgeon with lower expected ability, but whose ability she is certain of, over a surgeon with higher expected ability but for whom she is less certain about. The opposite could be true if physicians are willing to take a risk to try to get a good outcome.

The following proposition states the assumption under which the empirical results are consistent with Bayesian updating when the physician has risk preferences. It holds regardless of which group has higher variance but assume for now that the variance of beliefs for female surgeons' ability is larger than that for male surgeons.

Proposition 2. *If there are asymmetric changes in beliefs about male and female surgeons, it must be that the difference in the expected variance for female and male surgeons increases as the physician receives more signals:*

$$\sigma_w^2 - \sigma_m^2 < \mathbb{E}[\sigma_w^2|s] - \mathbb{E}[\sigma_m^2|s] \quad \forall s \in \mathcal{S}$$

Proof. See Appendix B. □

The proof in Appendix B shows that if $\sigma_w^2 - \sigma_m^2 < \mathbb{E}[\sigma_w^2|s] - \mathbb{E}[\sigma_m^2|s]$ does not hold, the physician must have a larger spread in beliefs about the group with the larger initial variance. Put differently, if the prior variance is larger for women, there is a larger spread in possible outcomes under female surgeons than male surgeons. This means that the physician should update more about women after *both* good and bad signals. Physicians will only update asymmetrically if the difference in posterior variance between women and men ($\mathbb{E}[\sigma_w^2|s] - \mathbb{E}[\sigma_m^2|s]$) is larger than the difference between the prior variances ($\sigma_w^2 - \sigma_m^2$) as the physician receives signals. This is a fairly strong assumption and does not hold for many of the distributions commonly used to model updating (for example, with normal or binomial distributions). Proposition 2 also holds for non-symmetric distributions and does not require the means of the distributions to be the same.

Discussion of Assumptions

The assumption that the difference in variance increases with signals is strong and there are few cases in which it would be violated. One such case is if good signals are informative for male surgeons but uninformative for female surgeons. In this case, the posterior for men could be much narrower than the posterior for women even if the prior on men is only slightly narrower than the prior on women, meaning that $\sigma_w^2 - \sigma_m^2 < \mathbb{E}[\sigma_w^2|s] - \mathbb{E}[\sigma_m^2|s]$ would hold. In Table 8, I test whether events are differentially predictive of future events for men and women and do not find any such evidence. However, it is possible that physicians hold these beliefs even if they do not match the data.

Distributions with odd higher-level moments can also violate the assumption.²⁶ Again, while I do not observe physicians' beliefs, I can measure the variance of surgeons' abilities. Table 1 gives the standard deviation of the ability measure for both men and women using the unmatched sample. The variance of women's abilities is slightly smaller than that of men. Figure 2 also shows that higher-level moments are roughly the same (there is no difference in skewness, for example). The difference in the variance of priors for male and female surgeons would thus have to be weakly increasing as the physician receives signals despite the actual variance in abilities being roughly the same. This again rejects rational expectations.

7.4 Alternative Models

Many existing discrimination models in labor economics start with the assumption that employers hold different beliefs about two groups²⁷ For example, if employers think that women are on average less skilled than men, they will be reluctant to hire women. However, after receiving signals from individual workers, employers correctly update based on these signals and can eventually end up with the same beliefs about a male and female worker. Once an employer holds the same beliefs about a man and a woman's abilities, the employer will form the same posterior about the man and the woman conditional on receiving the same signal from each.²⁸

My empirical results suggest that physicians may exhibit bias dynamically rather than through their priors. The surgeons I compare received the same number and fraction of

²⁶For example, if a physician has received few signals from women and the female ability distribution is strongly left-skewed relative to the male ability distribution, the observed asymmetries will appear.

²⁷For examples, see Coate and Loury (1993), Altonji and Pierret (2001), and Lange (2007)

²⁸There are also cases where beliefs do not converge to the worker's true ability, specifically if employees alter their behavior. Coate and Loury (1993), for example, show that potential workers may under-invest in skills depending on the standards that employers set. Glover et al. (2017) provide experimental evidence that grocery store clerks do change their behavior based on the degree of bias that managers exhibit.

a physician's referrals before an event. If referrals are a proxy of beliefs, the physicians should have the same priors about these surgeons, and the surgeons should be treated similarly after an event. Yet we see a male and a female surgeon for whom a physician holds the same prior being treated differently post-event.

Further, physicians' responses are strongest when they receive extreme or surprising signals. When the "placebo" surgeons produce the expected outcome, physicians do not treat men and women differently.

Several behavioral models exhibit agents who respond to signals in a biased way, either ignoring or misattributing signals depending on their priors.²⁹ Here, I discuss a simple example of physicians ignoring some signals, drawing on the psychology literature on attribution bias. I do not set up the model in full, but rather use features of attribution bias to show how it can lead to asymmetric updating.

Attribution Bias

Attribution theory studies how agents use and understand information to arrive at a conclusion, especially regarding an unexpected event (Fiske and Taylor, 1991). Typically, agents attribute information either to internal or external factors. Attribution bias occurs when there are systematic differences in how agents interpret information. For example, in the medical setting, physicians can attribute patient outcomes to the surgeon's ability (an internal factor) or to noise (an external factor). A physician exhibits attribution bias if she systematically interprets unexpectedly good or bad events differently depending on the surgeon's gender.

Under attribution bias, agents are unbiased when the actual outcome matches their expected outcome. However, if the outcome is far from an agent's expected outcome, the agent needs to decide whether her beliefs were wrong or whether the outcome was an anomaly. In doing so, the agent, either consciously or unconsciously, relies on stereotypes or deep-seated biases. For example, physicians might generally treat male and female surgeons equally, but hold a deep-seated belief that men are better surgeons. When a physician receives a signal far from her prior, she rationalizes it using this stereotype. Here, I briefly show how attribution bias produces asymmetries in physicians' responses to signals.

²⁹Confirmation bias, for example, assumes that agents misread signals to fit with their prior (Rabin and Schrag, 1999; Fryer et al., 2017). It assumes, though, that an agent's bias depends on her belief about the individual. I find that physicians update differently about two surgeons for whom they hold the same prior. Further, physicians should misread all signals to confirm their prior about an individual surgeon. However, I do not find evidence that physicians update after mundane events, such as general surgeries going as expected, as evidenced by the lack of change in referrals to placebo surgeons.

Updating about the Individual

As in the Bayesian case, physicians have the prior probability distribution $a_{i,t} \sim f(\bar{a}_i, \sigma_i^2)$ over a surgeon's ability. For simplicity, I consider the case in which physicians only care about a surgeon's mean ability. Physicians thus choose a surgeon $i \in I_j$ to maximize patient utility:

$$\operatorname{argmax}_{i \in I_j} U_p(a) = \beta \mathbb{E}[a_{i,t} | \bar{a}_i, g, s] + \epsilon_{ipt} \quad (15)$$

All variables are defined as before but I consider the case in which the physician sees a continuous signal, $s \sim f$. Physicians have an expected outcome of each patient referral, $\mathbb{E}[s | a_i]$.

When an actual outcome matches the physician's expected outcome, physicians update as Bayesians, relying on their prior about the individual surgeon to update. Specifically, the physician's posterior when updating as a Bayesian, denoted $f^b(a_i | s)$, is

$$f^b(a_i | s) \propto f(s | a_i) \cdot f(a_i)$$

If a physician has the same belief about a man and a woman's ability, the physician will update about them in the same way (e.g. form the same posterior about each) conditional on s . Empirically, this is what is seen in the "placebo" surgeon outcomes, where the surgical outcome is the expected outcome.

When the actual outcome does not match the expected outcome, physicians rationalize it, attributing the outcome to the surgeon's ability or to noise. If the physician attributes the outcome to the surgeon's ability, she updates as a Bayesian. If the physician attributes the outcome to noise, she throws away the signal and does not update. Assume that even if a physician believes that a given male and female surgeon are equally qualified, there is a societal stereotype that women make worse surgeons. As such, the physician attributes unexpectedly good signals ($s > \mathbb{E}[s]$) to ability if the surgeon is male. If the surgeon is female, however, the physician attributes it to noise with some probability q . Similarly, the physician attributes unexpectedly bad signals ($s < \mathbb{E}[s]$) to ability when the surgeon is female but attributes it to noise with probability q when the surgeon is male.

Compare this with a physician who exhibits attribution bias. Because the physician only accounts for a fraction q of a female surgeon's unexpectedly good outcomes and a fraction $1 - q$ of a male surgeon's unexpectedly bad outcomes, her posterior will be

distributed

$$f^c(a_i | s > \mathbb{E}[s]) \propto [1 - (1 - q) \cdot \mathbb{1}(g = w)] \cdot f^b(a_i | s) + (1 - q) \cdot \mathbb{1}(g = w) f(a_i) \quad (16)$$

$$f^c(a_i | s < \mathbb{E}[s]) \propto [1 - q \cdot \mathbb{1}(g = m)] \cdot f^b(a_i | s) + q \cdot \mathbb{1}(g = m) f(a_i) \quad (17)$$

after a good and bad signal respectively. In the above equations, $\mathbb{1}(g = w)$ is an indicator taking the value one if the surgeon is female. Similarly, $\mathbb{1}(g = m)$ takes the value one when the surgeon is male.

To help interpret these posteriors, consider a female surgeon. In equation 16, $\mathbb{1}(g = w) = 1$ and in equation 17, $\mathbb{1}(g = m) = 0$. If a female surgeon sends a good signal, the physician updates as a Bayesian with probability q and does not update with probability $1 - q$. In the latter case, the physician's beliefs do not change from the prior period. The physician's posterior is thus a weighted average of the Bayesian posterior and the physician's prior. However, if a male surgeon sends a good signal, the physician updates as a Bayesian. After a bad signal we see the opposite: the physician updates as a Bayesian if the surgeon is female and has a posterior that is a weighted average if the surgeon is male, producing the asymmetry observed in the data.

Updating about the Group

Attribution bias also helps explain the group-level results. Assume that physicians update iteratively about groups after seeing signals from different surgeons. Because women are underrepresented among surgeons, physicians will see fewer patient outcomes under women each quarter even if they are referring to them at their population share. If the variance of a_i is decreasing in the number of outcomes a physician sees, a common assumption, physicians' beliefs will be more diffuse for women than for men. That is, $\sigma_{i,m} < \sigma_{i,w}$, leading physicians to react more strongly to patient outcomes when the surgeon is female. However, if physicians exhibit attribution bias, the distribution about women's ability will shift more after bad outcomes than after good outcomes as physicians attribute female surgeons' good outcomes to noise with some probability. A physician's beliefs about a group can change, but it will take many more signals than it would if the physician was Bayesian.

8 Conclusion

Gender gaps in hiring, promotion, and pay persist in many industries. This paper identifies a mechanism that contributes to these gaps. Using referral volume to proxy for a

physician's beliefs about a surgeon's ability, I show that physicians exhibit asymmetric updating, lowering their referrals more to women than to men after bad outcomes and increasing them more to men than to women after good outcomes. In addition, physicians use their experience with one woman to infer the ability of other female surgeons. After a bad experience with one female surgeon, physicians become less likely to form new referral connections with women.

The results are consistent with Bayesian updating under specific assumptions on physicians' priors about surgeon ability. Physicians would have to believe that women are higher ability than men or that the variance of men and women's abilities shrinks differentially following signals. However the set of priors required do not match the actual distribution of surgeon ability. Therefore, although the results are reconcilable with Bayesian updating, they are inconsistent with physicians having rational expectations. Behavioral models in which physicians selectively ignore some signals, such as attribution bias, help to explain the empirical patterns.

Regardless of the model, the results have implications for how we think about employment decisions. Bayesian updating suggests that two individuals of equal ability and who perform equally well on a set of tasks will be evaluated in the same way. That is, an employer will hold the same posterior about each individual conditional on having the same prior and seeing the same signal from each. However, if employers exhibit asymmetric updating, two employees with the same objective performance could end up on different career tracks as the employer treats their performances differently. In this setting, women are punished for one mistake in the same way that men are punished for three mistakes, leading to lower skill accumulation and pay.

The implications of these results are especially important in occupations in which women are underrepresented. Because employers see relatively few signals from women, the performance of one woman influences the employers' beliefs about other women. If employers update asymmetrically, they might be too quick to let a woman go after a mistake and then be unwilling to hire more women in the future, preventing them from learning about the true distribution of women's abilities.

References

- [1] Altonji, Joseph and Charles Pierret. 2001. "Employer Learning and Statistical Discrimination." *The Quarterly Journal of Economics*, 116(1): 313-350.
- [2] Antecol, Heather, Kelly Bedard, and Jenna Stearns. 2016. "Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies?" IZA Discussion Paper No. 9904.
- [3] Arrow, Kenneth. 1973. "The Theory of Discrimination." In Orley Ashenfelter and Albert Rees, eds. *Discrimination in Labor Markets*. Princeton: Princeton University Press.
- [4] Azmat, Ghazala and Rosa Ferrer. 2017. "Gender Gaps in Performance: Evidence from Young Lawyers." *Journal of Political Economy*, 125(5): 1306-1355.
- [5] Babcock, Linda and Sara Laschever. 2004. *Women Don't Ask*. Princeton University Press.
- [6] Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva. 2017. "Does the Gender Composition of Scientific Committees Matter?" *American Economic Review*, 107(4): 1207-38.
- [7] Barber, Brad and Terrance Odean. 2001. "Boys Will be Boys: Gender, Overconfidence, and Common Stock Investment." *The Quarterly Journal of Economics*, 116(1): 261-292.
- [8] Barnett, Michael, Nancy Keating, Nicholas Christakis, James O'Malley, and Bruce Landon. 2012a. "Reasons for Choice of Referral Physician Among Primary Care and Specialist Physicians." *Journal of General Internal Medicine*, 27(5): 506-512.
- [9] Barnett, Michael, Ziruir Song, and Bruce Landon. 2012b. "Trends in Physician Referrals in the United States, 1999-2009." *Archives of Internal Medicine*, 172(2): 163-170.
- [10] Becker, Gary. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press.
- [11] Bertrand, Marianne and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review*, 94(4): 991-1013.
- [12] Beyer, Sylvia. 1990. "Gender Differences in the Accuracy of Self-Evaluations of Performance." *Journal of Personality and Social Psychology*, 59: 960-970.

- [13] Blau, Francine and Lawrence Kahn. 2016. "The Gender Wage Gap: Extend, Trends, and Explanations." NBER Working Paper No. 21913.
- [14] Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. "Stereotypes." *The Quarterly Journal of Economics*, 131(4): 1753-1794.
- [15] Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais. "'Acting Wife': Marriage Market Incentives and Labor Market Investments." *American Economic Review*, forthcoming.
- [16] Casadevall, Arturo and Jo Handelsman. 2014. "The Presence of Female Conveners Correlates with a Higher Proportion of Female Speakers at Scientific Symposia." *mBio*, 5(1): e00846-13.
- [17] Card, David, Ana Rute Cardoso, and Patrick Kline. 2016. "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women." *The Quarterly Journal of Economics*, 131(2): 633-686.
- [18] Ceci, Stephen, Donna Ginther, Shulamit Kahn, and Wendy Williams. 2014. "Women in Academic Science: A Changing Landscape." *Psychological Science in the Public Interest*, 15(3): 1-67.
- [19] Choudhry, Nitesh K., Joshua M. Liao, and Allan S. Detsky. 2014. "Selecting a Specialist: Adding Evidence to the Clinical Practice of Making Referrals." *Journal of the American Medical Association*, 312(18): 1861-1862.
- [20] Coate, Stephen and Glenn Loury. 1993. "Will Affirmative Action Policies Eliminate Negative Stereotypes?" *American Economic Review*, 83(5): 1220-12042.
- [21] Eil, David and Justin Rao. 2011. "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself." *American Economic Journal: Microeconomics*, 3(2): 114-38.
- [22] Egan, Mark L., Gregor Matvos, and Amit Seru. 2017. "When Harry Fired Sally: The Double Standard in Punishing Misconduct." NBER Working Paper No. 23242.
- [23] Farber, Henry and Robert Gibbons. 1996. "Learning and Wage Dynamics." *The Quarterly Journal of Economics*, 111(4): 1007-1047.
- [24] Fiske, Susan and Shelley Taylor. 1991. *Social Cognition* (2nd ed.). New York: McGraw-Hill.

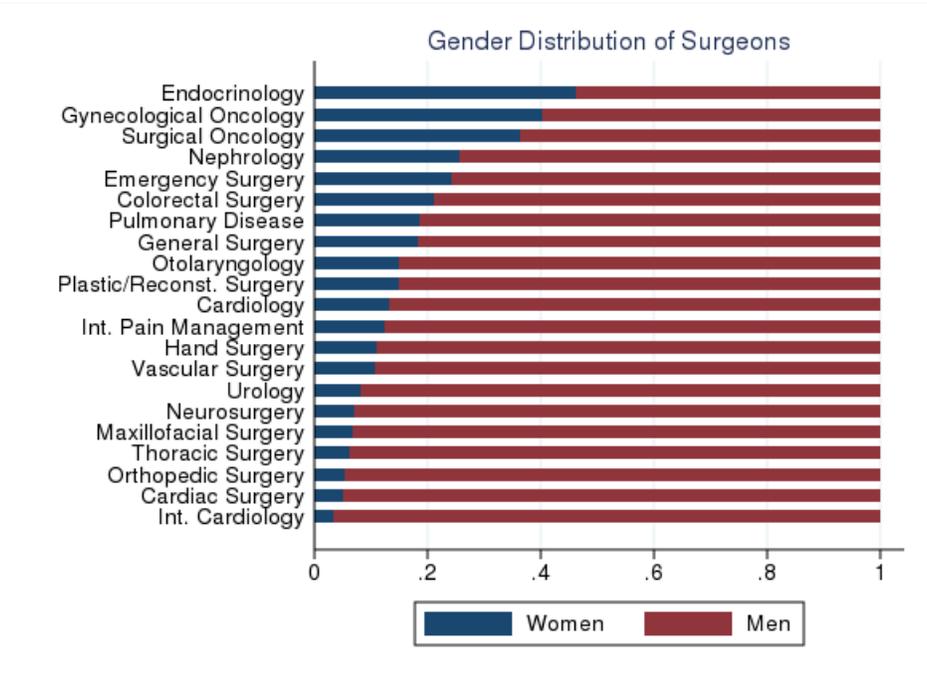
- [25] Forrest, Christopher, Paul Nutting, Sarah von Schrader, Charles Rohde, and Barbara Starfield. 2006. "Primary Care Physician Specialty Referral Decision Making: Patient, Physician, and Health Care System Determinants." *Medical Decision Making*, Jan-Feb.
- [26] Freedman, Rachel A., Elena M. Kouri, Dee W. West, and Nancy L. Keating. 2015. "Racial/Ethnic Differences in Patients' Selection of Surgeons and Hospitals for Breast Cancer Surgery." *JAMA Oncology*, 1(2): 222-230.
- [27] Fryer, Roland. 2007. "Belief Flipping in a Dynamic Model of Statistical Discrimination." *Journal of Public Economics*, 91(5-6): 1151-1166.
- [28] Fryer, Roland, Philipp Harms, and Matthew Jackson. 2017. "Updating Beliefs when Evidence is Open to Interpretation: Implications for Bias and Polarization." Working paper.
- [29] Ginther, Donna and Shulamit Kahn. 2004. "Women in Economics: Moving Up or Falling Off the Academic Career Ladder?" *Journal of Economics Perspectives*, 18(3): 193-214.
- [30] Glover, Dylan, Amanda Pallais, and William Pariente. 2017. "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores." *The Quarterly Journal of Economics*, 132(3): 1219-1260.
- [31] Goldin, Claudia. 2014a. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review*, 104(4): 1091-1119.
- [32] Goldin, Claudia. 2014b. "A Pollution Theory of Discrimination: Male and Female Differences in Occupations and Earnings." in L. Boustan, C. Frydman, and R. Margo, *Human Capital and History: The American Record* (Chicago: University of Chicago Press), pp. 313-48.
- [33] Goldin, Claudia, and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians," *American Economic Review*, 90(4): 715-741.
- [34] Gregory, A., R.J. Skiba, and P.A. Noguera. 2010. "The Achievement Gap and the Discipline Gap: Two Sides of the Same Coin?" *Educational Researcher*, 39: 59-68.
- [35] Hughes, Gordon. 1991. "The Learning Curve in Staples Surgery." *The Laryngoscope*, 101(12): 1280-1284.

- [36] Jena, Anupam, Andrew Olenski, and Daniel Blumental. 2016. "Sex Differences in Physician Salary in US Public Medical Schools." *JAMA Internal Medicine*, 176(9): 1294-1304.
- [37] Kahn, Lisa and Fabian Lange. 2014. "Employer Learning, Productivity, and the Earnings Distribution: Evidence from Performance Measures." *The Review of Economic Studies*, 81(4): 1575-1613.
- [38] Keating, Nancy L., A. James O'Malley, Jukka-Pekka Onnela, and Bruce E. Landon. 2017. "Assessing the Impact of Colonoscopy Complications on Use of Colonoscopy Among Primary Care Physicians and Other Connected Physicians: An Observational Study of Older Americans." *BMJ Open*, 7.
- [39] Keehner, Madeleine, Yvonne Lippa, Daniel Montello, Frank Tendick, and Mary Hegarty. 2006. "Learning a Spatial Skill for Surgery: How the Contributions of Abilities Change with Practice." *Applied Cognitive Psychology*, 20(4): 487-503.
- [40] Kinchen, Kraig, Lisa Cooper, David Levine, Nae Yuh Wang, and Neil Powe. 2004. "Referral of Patients to Specialists: Factors Affecting Choice of Specialist by Primary Care Physicians." *Annals of Family Medicine*, 2(3): 245-252.
- [41] Kleven, Henrik, Camille Landais, and Jakob Egholt Sogaard. 2017. "Children and Gender Inequality: Evidence from Denmark." Working paper.
- [42] Lautenberger, Diana, Valerie Dandar, Claudia Raezer, and Rae Anne Sloane. 204. "The State of Women in Academic Medicine." Association of American Medical Colleges Annual Report.
- [43] Lichtenstein, S., B. Fischhoff, and L.D. Phillips. 1982. Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases*, (pp. 306-334). Cambridge, England: Cambridge University Press.
- [44] Lo Sasso, Anthony, Michael Richards, Chiu-Fang Chou, and Susan Gerber. 2011. "The \$16,819 Pay Gap for Newly Trained Physicians: The Unexplained Trend of Men Earning More Than Women." *Health Affairs*, 30(2): 193-201.
- [45] Ly, Dan, Seth Seabury, and Anupam Jena. 2016. "Differences in Incomes of Physicians in the United States by Race and Sex: Observational Study." *BMJ*, 353: i2921.

- [46] Medscape Physician Compensation Report. 2017. Sarah Grisham, Senior Editor. Art Science Code LLC. <<http://www.medscape.com/slideshow/compensation-2017-overview-6008547>>
- [47] Mobius, Markus, Muriel Niederle, Paul Niehaus, and Tanya Rosenblat. 2015. "Managing Self-Confidence." Working paper.
- [48] Niederle, Muriel and Lise Vesterlund. 2007. "Do Women Shy Away From Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics*, 122(3): 1067-1101.
- [49] Noonan, Mary, Mary Corcoran, and Paul Courant. 2005. "Pay Differences Among the Highly Trained: Cohort Differences in the Sex Gap in Lawyers' Earnings." *Social Forces*, 84(2): 851-870.
- [50] Phelps, Edmond. "The Statistical Theory of Racism and Sexism." *American Economic Review*, 62(4): 659-661.
- [51] Pinkston, Joshua. 2009. "A Model of Asymmetric Employer Learning with Testable Implications." *The Review of Economic Studies*, 76(1): 367-394.
- [52] Rabin, Matthew and Joel Schragg. 1999. "First Impressions Matter: A Model of Confirmatory Bias." *Quarterly Journal of Economics*, 114(1): 37-82.
- [53] Sasser, Alicia. 2005. "Gender Differences in Physician Pay: Tradeoffs Between Career and Family." *The Journal of Human Resources*, 40(2): 477-504.
- [54] Shea, Dennis, Bruce Stuart, Joseph Vasey, and Soma Ng. 1999. "Medicare Physician Referral Patterns." *Health Services Research*, 34(1): 331-348.
- [55] Skiba, Russell J., Robert S. Michael, Abra Carroll Nardo, and Reece L. Peterson. 2002. "The Color of Discipline: Sources of Racial and Gender Disproportionality in School Punishment." *The Urban Review*, 34(3): 317-342.
- [56] Zeltzer, Dan. 2017. "Gender Homophily in Referral Networks: Consequences for the Medicare Physician Earnings Gap." Working paper. Available at SSRN: <https://ssrn.com/abstract=2921482>

Figures

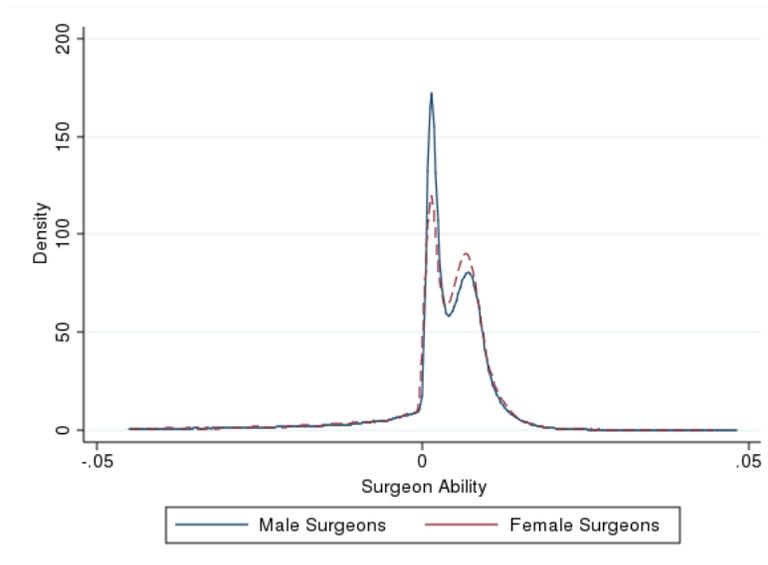
FIGURE 1: GENDER BREAKDOWN OF SPECIALTIES



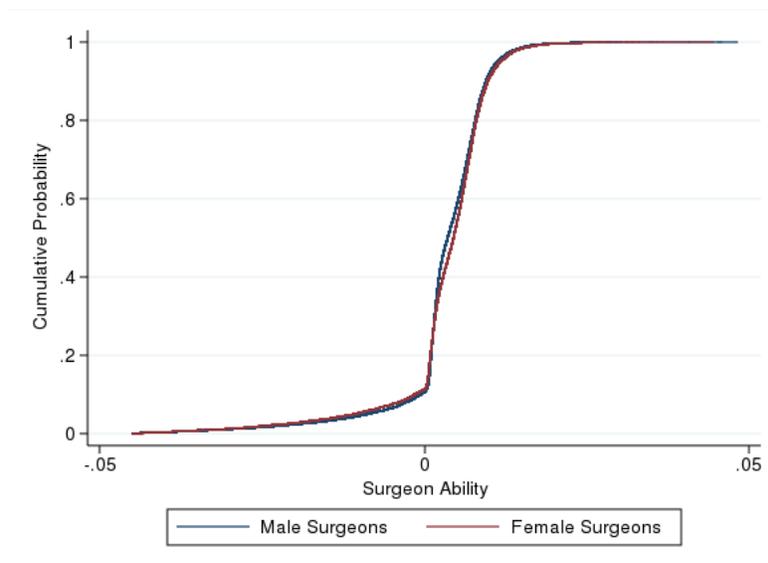
Notes: This figure displays the surgical specialties that could be included in the analysis (e.g. the specialties available to match on) and the gender composition of each. The gender distributions are calculated using the number of surgeons represented in each specialty in the Medicare dataset and therefore may be slightly different than the gender distribution when including all surgeons (such as those not accepting Medicare). Int. Cardiology is Internal Cardiology and Int. Pain Management is Internal Pain Management.

FIGURE 2: DISTRIBUTION OF SURGEON ABILITY

(a) Ability PDF

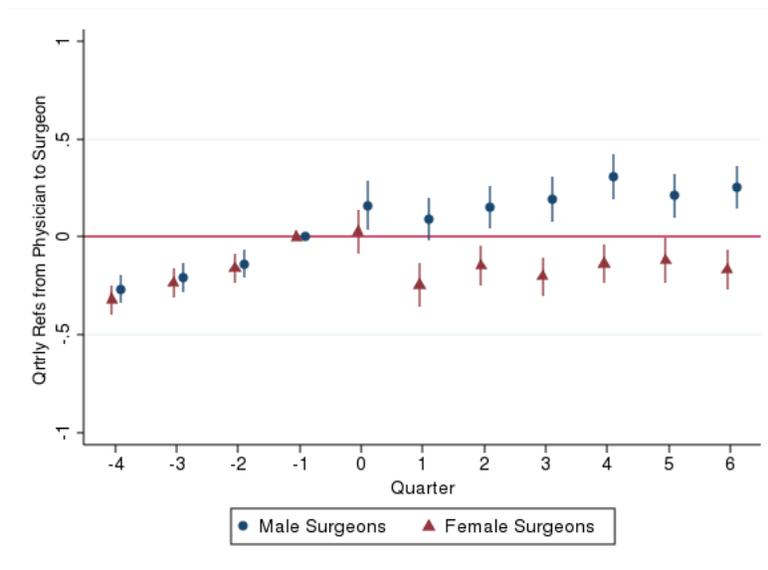


(b) Ability CDF



Notes: This figure shows the distributions of male and female surgeon ability, where ability is defined as described in Section 2.2.1. The top figure is the PDF and the bottom figure is the CDF. The full sample of surgical specialists is used to create these distributions.

FIGURE 3: QUARTERLY ESTIMATES OF PHYSICIAN'S REACTION TO DEATH

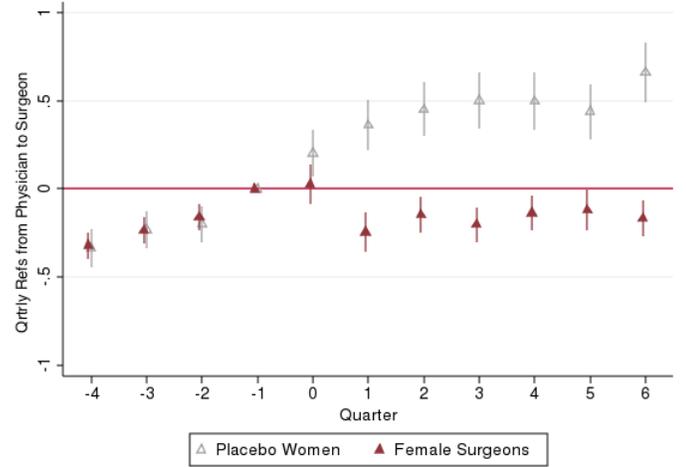
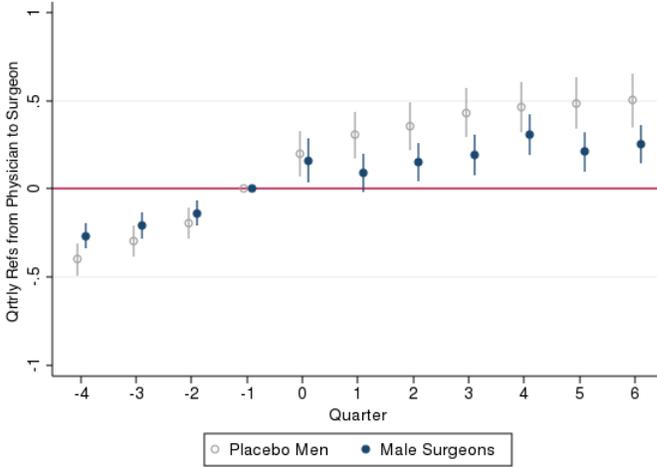


Notes: This figure plots the quarterly regression coefficients and 95% confidence intervals from estimating equation 3 using the sample of matched male and female surgeons who experience a patient death. The coefficients are plotted relative to the number of referrals the surgeon was receiving from the physician in $k = -1$, which are normalized to zero. In $k = -1$, male and female surgeons both received an average of 0.65 referrals from the referring physician j . A patient that physician j referred to surgeon i dies in $k = 0$. The outcome variable is the total number of referrals that physician j sends to surgeon i each quarter. Standard errors are clustered at the physician-surgeon match level.

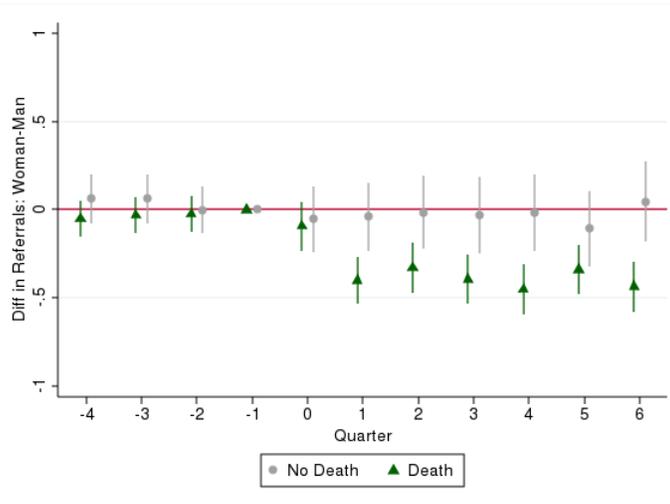
FIGURE 4: COMPARISON WITH PLACEBO SURGEONS

(a) Male Surgeons

(b) Female Surgeons



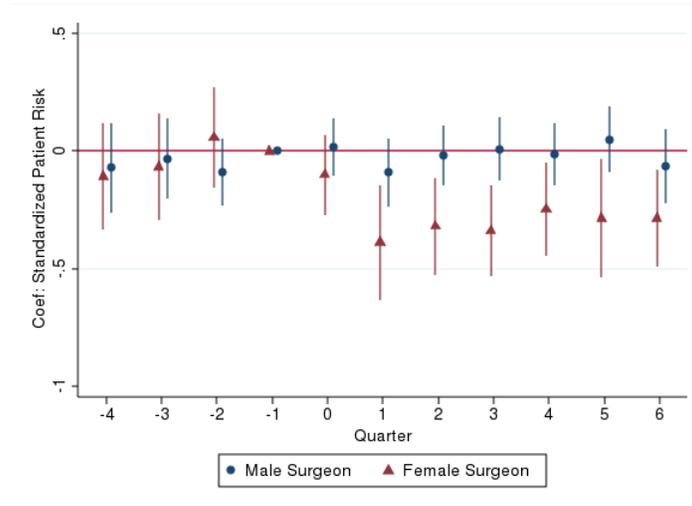
(c) Difference-in-Differences



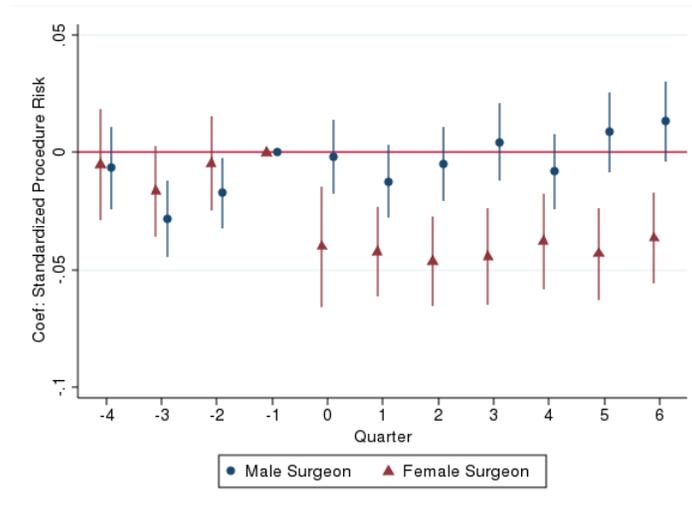
Notes: This figure shows three plots that compare the impact of a bad event to “placebo” outcomes. Panels (a) and (b) plot the quarterly regression coefficients and 95% confidence intervals from estimating equation 4 on the matched sample of male (panel a) and female (panel b) surgeons who did not experience a patient death. These coefficients are represented by the hollow circles and triangles. For the placebo outcomes in each panel, physician j refers a patient to surgeon i and the surgery does not result in a patient death. I then plot the quarterly coefficients from Figure 3. Panel (c) shows difference-in-differences estimates and 95% confidence intervals. The estimates show the difference between referrals to a female and a male surgeon in quarter k . The grey circles show the estimates for the sample of placebo surgeons and the green triangles show the estimates for the sample of matched surgeons who experience a patient death. All coefficients are plotted relative to the number of referrals the surgeon was receiving from the physician in $k = -1$, which are normalized to zero. Standard errors are clustered at the physician-surgeon match level.

FIGURE 5: RISKINESS OF FUTURE PROCEDURES AND PATIENTS

(a) Patient Risk



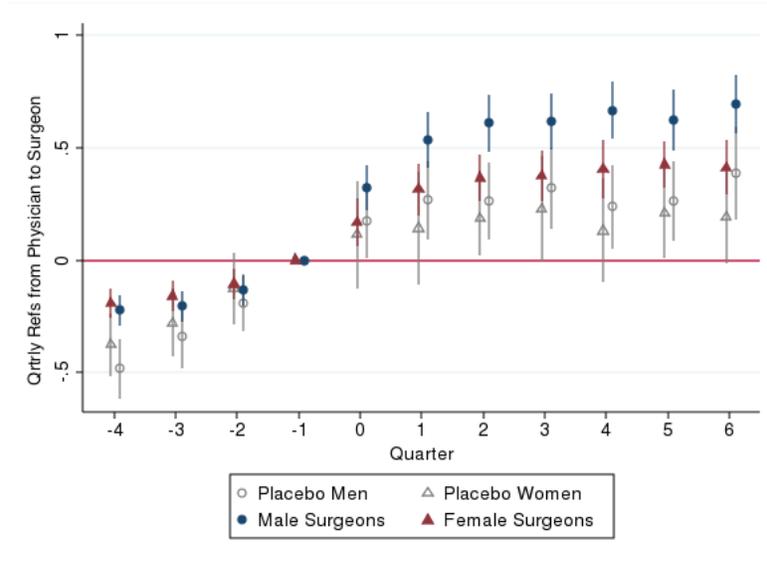
(b) Procedure Risk



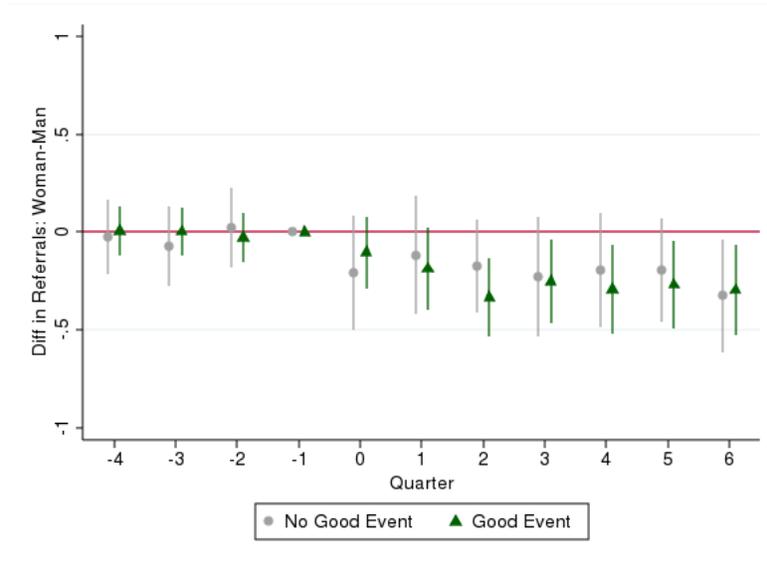
Notes: Panels (a) and (b) show how the riskiness of future patients and procedures change after a patient death. The outcome variable in Panel (a) is the average patient risk of the patients that physician j sends to surgeon i in quarter k , excluding the risk level of the patient who died. The outcome variable in Panel (b) is the average death rate of the procedures that physician j sends to surgeon i in quarter k , again excluding the riskiness of the procedure that the patient who died was referred for. The coefficients are plotted relative to the average patient risk or procedure death rate that the surgeon received in $k = -1$, which is normalized to zero. Standard errors are clustered at the physician-surgeon match level. See Section 2 for the definition and calculation of patient risk.

FIGURE 6: QUARTERLY ESTIMATES FOR UNEXPECTEDLY GOOD OUTCOMES

(a) Treated and Placebo Surgeons

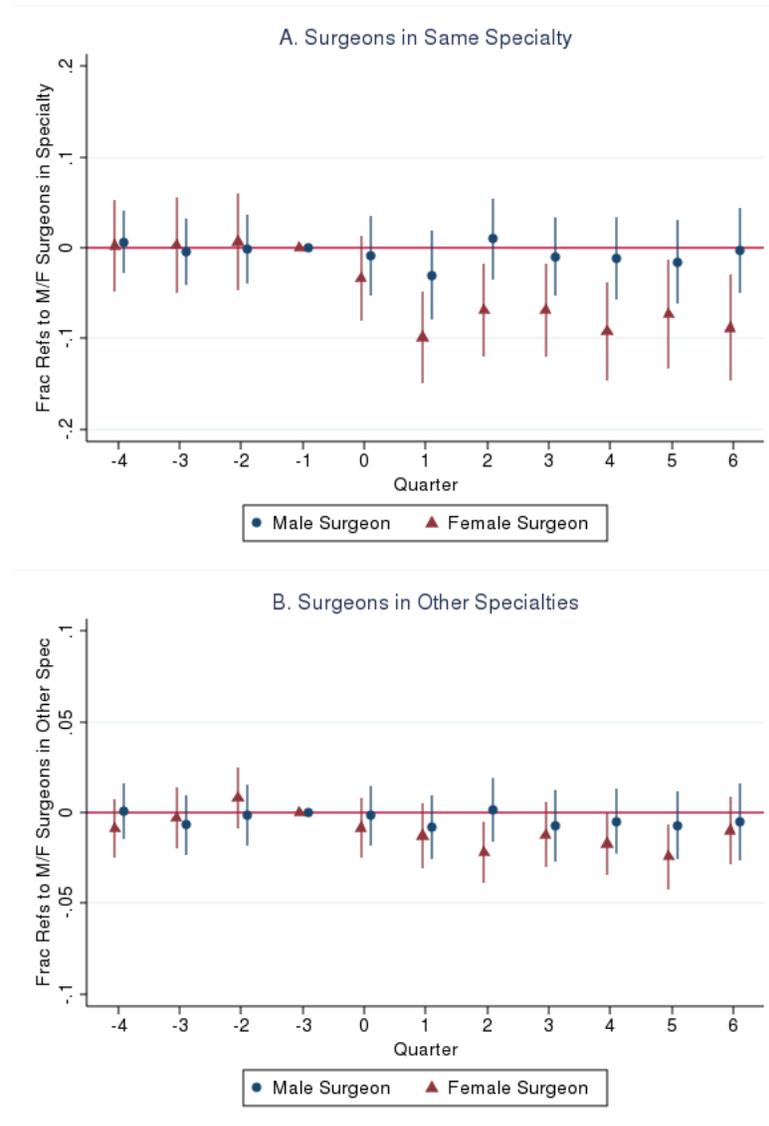


(b) Difference-in-Differences



Notes: Panel (a) in this figure shows the quarterly coefficients and 95% confidence intervals from estimating equations 3 and 4. The blue circles and red triangles show the impact that an unanticipated patient survival has on referrals to the male and female surgeon respectively. The grey circles and triangles show what would have happened in the absence of a survival. Panel (b) shows the difference-in-differences estimates and corresponding 95% confidence intervals. The grey circles show the estimates for the sample of placebo surgeons and the green triangles show the estimates for the sample of matched surgeons who have an unanticipated patient survival. All coefficients are plotted relative to the number of referrals the surgeon was receiving from the physician in $k = -1$, which are normalized to zero. Standard errors are clustered at the physician-surgeon match level.

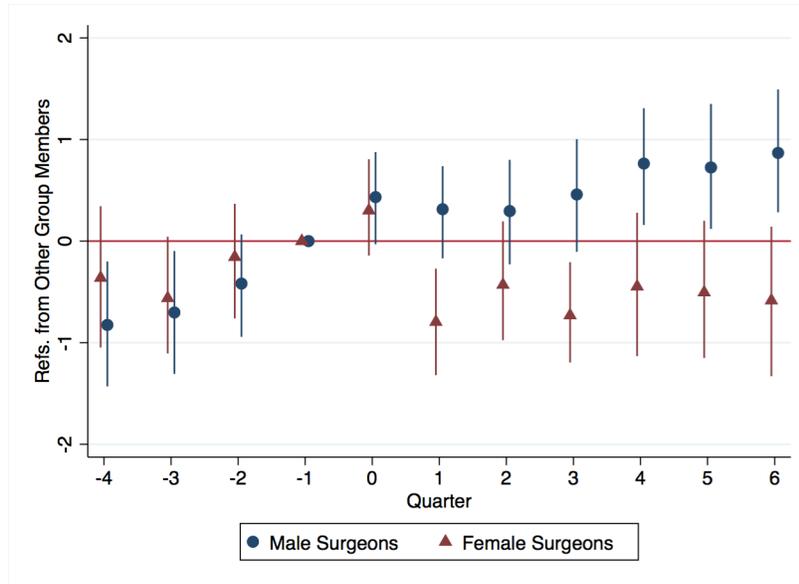
FIGURE 7: SPILLOVERS TO OTHER SURGEONS AFTER BAD OUTCOME



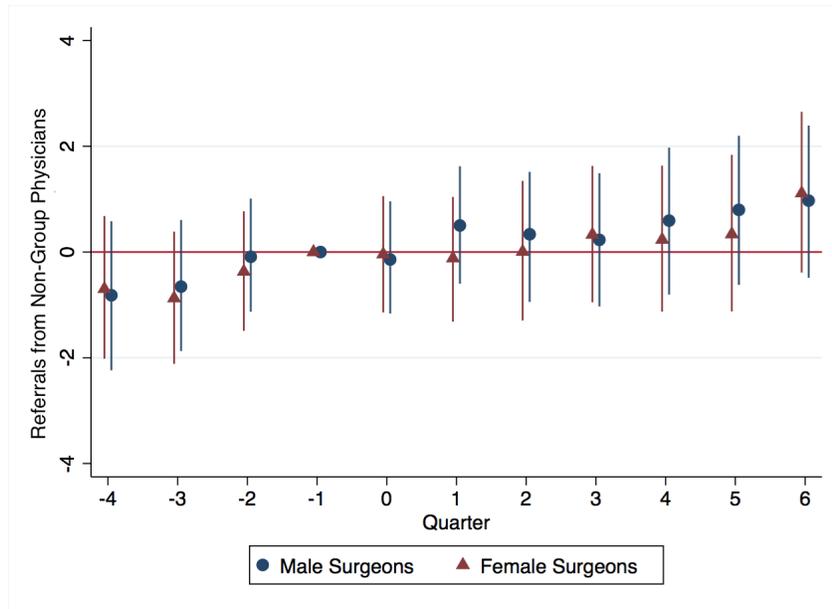
Notes: This figure shows how physicians change their behaviour toward other surgeons after a patient death. In $k = 0$, a patient that physician j sent to surgeon i dies. The outcome variable is the fraction of new referrals going to female or male surgeons (whom physician j hasn't referred to before) in the same specialty as surgeon i (Panel A) or a different specialty (Panel B). Both outcomes variables exclude the performing surgeon. The blue circles are estimated from a regression using the fraction of referrals going to men as the outcome and the red triangles are estimated from a regression using the fraction of referrals going to women as the outcome. The coefficients are plotted relative to the number of referrals that physician j was sending to surgeons she had not previously referred to in quarter $k = -1$. I control for the fraction of available surgeon who are male or female and also include physician-surgeon match fixed effects. Standard errors are clustered at the physician-surgeon match level.

FIGURE 8: INFORMATION SPILLOVERS TO OTHER PHYSICIANS

(a) Referrals from Physicians in Group Practice

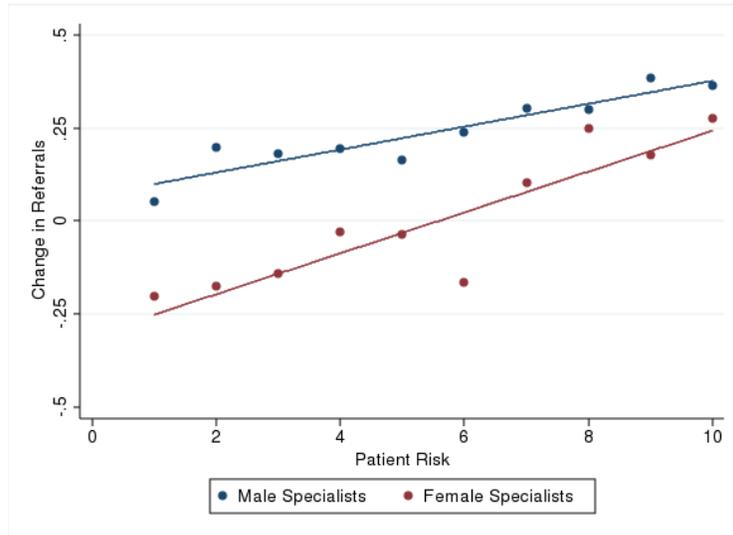


(b) Referrals from Physicians Outside of Practice



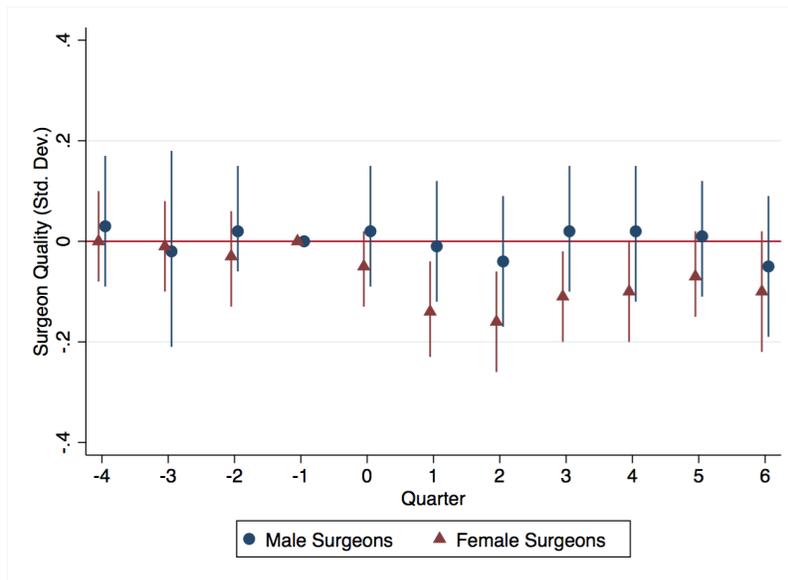
Notes: This figure shows how referrals from other physicians to a performing surgeon change after a bad event. The outcome variable in Panel (a) is the number of quarterly referrals from physicians in the referring physician’s group practice. The outcome variable in Panel (b) is the number of quarterly referrals from physicians outside of the referring physician’s group practice but in the same Hospital Referral Region. Both panels are estimated on the sample of matched male and female surgeons who experience a patient death. All coefficients are plotted relative to the number of referrals the surgeon was receiving from the physician in $k = -1$, which are normalized to zero. Standard errors are clustered at the physician-surgeon match level.

FIGURE 9: PHYSICIAN RESPONSE BY PATIENT RISK



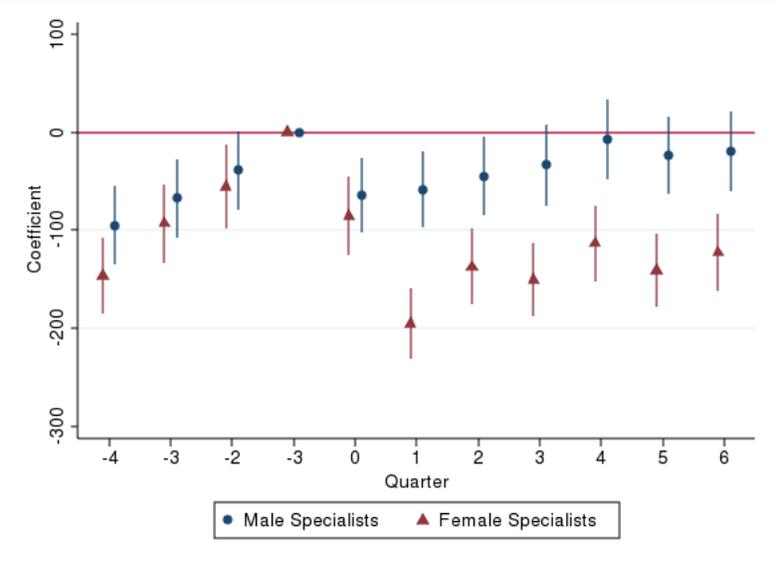
Notes: This binned scatterplot shows the relationship between the risk of the patient that died and the change in referrals from physician j to surgeon i , using the matched sample of male and female surgeons who experience a patient death. Patient risk is calculated as in Section 2 and is then binned into deciles. Both variables are residualized on a physician-surgeon match pair effect. The line of best fit using OLS is shown separately for male and female surgeons. The lines of best fit have slopes of 0.060 (s.e. = 0.004) for female surgeon and 0.026 (s.e. = 0.002) for male surgeons.

FIGURE 10: CHANGE IN SURGEON QUALITY



Notes: This figure shows how the quality of surgeon that a physician refers to changes after a patient death. In the figure, a patient that surgeon i received from physician j dies in quarter $k = 0$. The outcome variable is the standard deviation of surgeon ability from the mean. The coefficients are plotted relative to the standard deviation of surgeon quality that the physician was referring to in $k = -1$. For example, if a physician was referring to surgeons who were an average of 1 standard deviation above the mean in $k = -1$, all coefficients are plotted relative to 1. Standard errors are clustered at the physician level.

FIGURE 11: MEDICARE PAYMENTS



Notes: This figure shows plots quarterly Medicare payments using the sample of matched surgeons who experience a patient death. The outcome variable is the total quarterly Medicare payments that surgeon i received from j 's referrals. The coefficients are plotted relative to the total Medicare payments that the surgeon was receiving from the physician in $k = -1$. Payments in $k = -1$ are normalized to zero. A patient that physician j referred to surgeon i dies in $k = 0$. Standard errors are clustered at the physician-surgeon match level.

Tables

TABLE 1: SUMMARY STATISTICS: UNMATCHED SAMPLE

	Male Surgeons		Female Surgeons		p-value
	Mean	SD	Mean	SD	
<i>Panel A: Surgeons</i>					
Total Patients	112.4	170.0	81.2	144.1	0.001
Freq. of Bad Events (%)	0.85	4.4	0.72	4.3	0.001
Freq. of Good Events (%)	26.0	27.0	26.1	26.0	0.372
# Physicians Referring	18.6	27.1	14.7	25.2	0.001
Patient Risk (%)	0.48	0.35	0.40	0.33	0.001
Patient Female (%)	52.9	30.0	58.5	32.5	0.001
Patient Minority (%)	17.6	25.4	18.5	28.0	0.001
Patient Age	71.8	8.3	70.8	9.7	0.001
Surgeon Ability	0.0011	0.0053	0.0012	0.0052	0.001
Observations	152,237		30,603		
<i>Panel B: Referring Physicians</i>					
	Male Surgeons		Female Surgeons		p-value
	Mean	SD	Mean	SD	
# Surgeons Referred To	15.5	13.2	2.5	3.2	0.001
# Surgeons Referred to for a Proc.	1.41	1.24	0.17	0.51	0.001
Observations	181,237		50,603		

Notes: This table shows summary statistics for the full sample of surgeons. The variables are measured at the yearly so that Total Patients is the average number of patients a surgeon receives in a year, for example. Total Events is the number of good and bad events a surgeon experiences and # of Physicians Referring is the average number of physicians that a given surgeon receives referrals from in a year. Patient Risk and Surgeon Ability are calculated as in Section 2

TABLE 2: BALANCE TABLE: MATCHED SAMPLE FOR BAD OUTCOMES

	Male Surgeons		Female Surgeons		p-value
	Mean	SD	Mean	SD	
Patient Refs from Physician	24.9	33.7	25.2	34.3	0.639
Physician's Refs for Proc. (%)	50.6	14.0	50.7	14.1	0.454
Total Patient Refs	74.2	116.5	73.4	115.2	0.651
Patient Age	77.2	12.0	77.2	12.1	0.887
Patient Minority (%)	8.4	27.7	8.4	27.7	0.999
Patient Female (%)	49.6	50.0	49.6	50.0	0.999
Patient Risk (%)	0.81	0.40	0.81	0.40	0.915
Risk All Past Ptnts	0.007	0.004	0.007	0.004	0.830
Total Procs. Performed	10.5	15.5	10.6	15.5	0.593
Years of Experience	23.3	8.1	23.2	8.0	0.174
Available Surgeons	44.5	27.8	45.1	27.8	0.137
# of Matched Surgeons	7,757		7,757		
# Unique Surgeons	5,579		3,561		

Notes: The balance table shows summary statistics for the matched sample of surgeons who experience a death. The sample is matched exactly on patient gender and minority status as well as surgeon specialty and procedure (not shown above). The sample is matched coarsely on all other variables. Patient Refs from Physician is the number of referrals that surgeon i received from physician j in quarters $t = -5$ through $t = -2$ where patient p dies in $t = 0$. Physician's Refs for Proc. is the percent of physician j 's referrals for the procedure being performed that went to surgeon i before the event. Similarly, Total Patient Refs is the number of patients surgeon i received from any physician during this period. Patient Risk is the risk level of the patient who dies while Risk All Past Ptnts is the average risk of all past patients surgeon i received between $t = -5$ and $t = -2$. Total Procs. Performed is the number times surgeon i performed the procedure that physician j refers for between $t = -5$ and $t = -2$. Years of Experience is the number of years since the surgeon graduated from medical school. Available Surgeons is the number of surgeons who are in the same specialty as surgeon i that the referring physician has the option of referring to. This is measured using Hospital Referral Region as the geographical unit.

TABLE 3: BALANCE TABLE: MATCHED SAMPLE FOR GOOD OUTCOMES

	Male Surgeons		Female Surgeons		p-value
	Mean	SD	Mean	SD	
Patient Refs from Physician	15.5	21.0	15.5	21.0	0.481
Physician's Refs for Proc. (%)	34.2	7.3	34.1	7.1	0.835
Total Patient Refs	45.5	68.5	45.0	68.0	0.645
Patient Age	91.3	4.7	91.3	4.6	0.353
Patient Minority (%)	4.8	21.3	4.8	21.3	0.999
Patient Female (%)	24.6	43.1	24.6	43.1	0.999
Patient Risk	0.017	0.002	0.017	0.002	0.650
Risk All Past Ptnts	0.015	0.004	0.015	0.004	0.850
Total Procs. Performed	8.0	15.5	8.0	15.3	0.861
Years of Experience	22.9	8.4	22.7	8.5	0.461
Available Surgeons	38.9	25.9	38.7	25.7	0.603
# Matched Surgeons	6,979		6,979		
# Unique Surgeons	5,554		2,882		

Notes: The balance table shows summary statistics for the matched sample of surgeons who experience an “unexpectedly good” outcome. The sample is matched exactly on patient gender and minority status as well as surgeon specialty and procedure (not shown above). The sample is matched coarsely on all other variables. Patient Refs from Physician is the number of referrals that surgeon i received from physician j in quarters $t = -5$ through $t = -2$ where patient p dies in $t = 0$. Physician's Refs for Proc. is the percent of physician j 's referrals for the procedure being performed that went to surgeon i before the event. Similarly, Total Patients Refs is the number of patients surgeon i received from any physician during this period. Patient Risk is the risk level of the patient who dies while Risk All Past Ptnts is the average risk of all past patients surgeon i received between $t = -5$ and $t = -2$. Total Procs. Performed is the number times surgeon i performed the procedure that physician j refers for between $t = -5$ and $t = -2$. Years of Experience is the number of years since the surgeon graduated from medical school. Available Surgeons is the number of surgeons who are in the same specialty as surgeon i that the referring physician has the option of referring to. This is measured using Hospital Referral Region as the geographical unit.

TABLE 4: IMPACT OF EVENT ON REFERRALS TO PERFORMING SURGEON

Event	Referrals		Medicare Pay (\$)	
	(1) Bad	(2) Good	(3) Bad	(4) Good
Post	0.006 (0.058)	0.509*** (0.069)	-80.05*** (25.09)	145.29*** (45.21)
Female × Post	-0.291*** (0.087)	-0.222** (0.100)	-138.97*** (32.54)	-22.88 (64.14)
Time Trend	0.099*** (0.011)	0.073*** (0.011)	31.84*** (5.84)	27.58*** (8.07)
Female × Time Trend	-0.009 (0.013)	-0.010 (0.015)	13.67*** (4.65)	-6.08 (8.59)
Post × Time Trend	-0.072*** (0.012)	-0.046*** (0.012)	-35.51*** (6.00)	-41.83*** (8.82)
<i>Average Post Effect On:</i>				
Male Surgeons	0.101	0.604	-92.90	95.42
Female Surgeons	-0.222	0.346	-184.02	51.26
Mean of Outcome Var.	0.65	0.48	309.17	264.86
Observations	34,053	29,214	34,053	29,214
Clusters	3,425	2,948	3,425	2,948
R-Squared	0.265	0.325	0.237	0.249

Notes: This table displays the effect of a bad event (columns 1 and 3) and a good event (columns 2 and 3) on referrals and Medicare payments to the performing surgeon. In columns 1 and 2, the outcome variable is the number of referrals from the referring physician to the performing surgeon in a quarter. In columns 3 and 4, the outcome variable is the total Medicare pay that the surgeon receives from referrals from the physician. All regressions include surgeon-physician match fixed effects and standard errors are clustered at the surgeon-physician match level. Columns 1 and 3 are estimated on the sample of matched male and female surgeons who experience a patient death. Columns 2 and 4 are estimated on the sample of matched male and female surgeons who experience a good patient outcome. Both samples are limited to physician-surgeon pairs in which the physician had referred between 1 and 10 patients in the past. Physician-surgeon match fixed effects are included in all regressions and standard errors are clustered at the physician-surgeon match level. Levels of significance: *10%, ** 5%, and *** 1% level.

TABLE 5: IMPACT OF AN EVENT ON REFERRALS TO OTHERS

Fraction of New Referrals to Women in:	Same Specialty		Other Specialty
	(1)	(2)	(3)
Event	Bad	Good	Bad
Post	0.011 (0.025)	-0.022 (0.028)	0.007 (0.012)
Female × Post	-0.097** (0.039)	0.034 (0.044)	-0.028 (0.018)
Time Trend	0.003 (0.006)	0.003 (0.008)	0.004 (0.003)
Female × Time Trend	-0.003 (0.010)	0.002 (0.012)	0.001 (0.004)
Post × Time Trend	-0.004 (0.007)	-0.001 (0.009)	-0.005 (0.004)
Female × Post × Time Trend	0.006 (0.012)	0.001 (0.014)	-0.001 (0.005)
Avg. Post Effect	-0.079	0.029	-0.025
Mean of Outcome Var.	0.15	0.12	0.13
Observations	34,053	29,214	34,053
Clusters	3,425	2,948	3,425
R-Squared	0.417	0.350	0.206

Notes: This table displays the effect of a bad (columns 1 and 3) or a good (column 2) event by a performing surgeon on referrals to other surgeons of the same gender. The outcome variable is the fraction of a physician's new referrals that go to men or women, excluding the performing surgeon. Columns 1 and 2 look at the fraction of new referrals going to women/men in the same specialty as the performing surgeon while Column 3 looks at the fraction of new referrals going to women/men in other specialties. I control for the fraction of available of surgeons who are women or men within the same specialty (Columns 1 and 2) or in any specialty (Columns 3). Regressions are estimated using the sample of matched male and female surgeons who experience a patient death or an unexpectedly good outcome. Levels of significance: *10%, ** 5%, and *** 1% level.

TABLE 6: EXPERIENCE WITH SURGEONS AND PHYSICIAN RESPONSE

Outcome Var:	Referrals from Physician to Performing Surgeon			
	(1)	(2)	(3)	(4)
	Bad Outcome		Good Outcome	
Post	-0.084 (0.061)	0.247*** (0.075)	0.379*** (0.064)	0.619*** (0.089)
Female × Post	-0.475*** (0.068)	-0.548*** (0.102)	-0.285*** (0.061)	-0.511*** (0.128)
Post × Pre-Referrals	0.014** (0.005)		-0.015** (0.007)	
Female × Post × Pre-Referrals	0.008* (0.004)		0.013** (0.006)	
Post × High Propensity		-0.145 (0.090)		-0.147 (0.135)
Female × Post × High Propensity		0.210** (0.103)		0.373** (0.159)
Mean of Outcome Var.	1.30	0.54	1.04	0.54
Observations	48,610	26,387	39,640	20,245
Clusters	4,972	2,378	4,025	1,745

Notes: This table looks at how a physician’s relationship with the performing surgeon and with other surgeons changes her response. The outcome variable is the number of patients that the referring physician sends to the performing surgeon in a quarter. “Pre-Referrals” is the number of referrals that surgeon i received from physician j before the patient death. “High Propensity” is an dummy variable that equals one when a physician sends a larger fraction of her referrals to female surgeons than the average physician in her HRR. All regressions include a time trend and time trend interactions (with Female Surgeon, Post Event and the Pre-Referrals/High Propensity) as well as physician-surgeon match fixed effects. Standard errors are clustered at the physician-surgeon match level. Levels of significance: *10%, ** 5%, and *** 1% level.

TABLE 7: VARIABLES CORRELATED WITH SURGEON RESPONSE

X-Var:	Referrals to Surgeon			
	Num. Surgeons (1)	Surgeon Exp. (2)	Physician Exp. (3)	Fem. Physician (4)
Post	-0.256* (0.129)	-0.043 (0.131)	0.177 (0.160)	0.109 (0.058)
Female × Post	-0.522*** (0.119)	-0.306** (0.130)	-0.518** (0.199)	-0.434*** (0.056)
Post × X-Var	0.004 (0.002)	0.002 (0.006)	-0.004 (0.006)	0.058 (0.140)
Female × Post × X-Var	0.004 (0.002)	0.003 (0.010)	0.008 (0.008)	0.185 (0.123)
Observations	17,691	34,054	28,989	28,989
Clusters	1,781	2,945	2,578	2,914

Notes: This table tests whether the four variables listed at the top of each column are correlated with a physician’s response. Each variable is interacted with the Post and Post×Female variables. “Num. Surgeons” is the number of surgeons in the same specialty as the performing surgeon who work in the physician’s Hospital Referral Region (i.e. the physician’s outside options). “Surgeon Exp.” and “Physician Exp.” are the surgeon and physician’s experience, measured as the number of years that each doctor has been out of medical school. “Physician Fem.” is a dummy variable indicating that the physician is a woman. A time trend, time trend interactions with all variables, and a physician-surgeon match fixed effect are included in each regression. Columns 2 and 3 also include surgeon experience squared and physician experience squared respectively. Standard errors are clustered at the physician-surgeon match level. Levels of significance: *10%, ** 5%, and *** 1% level.

TABLE 8: ARE EVENTS PREDICTIVE OF FUTURE EVENTS?

	# Bad Events (1)	Any Bad Event (2)	# Good Events (3)
Female Surgeon	-0.792** (0.366)	-0.031*** (0.011)	-0.045 (0.177)
Log Future Ptnt Risk	0.283 (0.197)	0.223*** (0.017)	0.859** (0.321)
Future Referrals	0.006*** (0.001)	0.001** (0.001)	0.005*** (0.001)
Log Past Ptnt Risk	-1.173*** (0.380)	-0.038*** (0.005)	0.243*** (0.070)
Past Referrals	-0.001* (0.001)	0.001*** (0.000)	-0.002*** (0.000)
Surgeon Experience	0.001 (0.013)	-0.001 (0.001)	0.012*** (0.004)
Observations	24,459	24,459	24,478
R-Squared	0.594	0.373	0.418

Notes: This table shows the results from testing whether events are differentially predictive of future events for male and female surgeons. The sample consists of the full sample of surgeons who every experienced a bad or good event. Log Future Patient Risk is the average risk of the future patients. Future Referrals is the number of referrals that the surgeon receives from any physician in the future. Similarly, # Past Referrals and Log Patient Past Risk are the number of referrals and average risk of those patients that the surgeon received before the patient death. Surgeon specialty fixed effects are also included. Standard errors are clustered at the specialty level. Levels of significance: *10%, ** 5%, and *** 1% level.

Appendix A Data Appendix

A.1 Example of Procedure Code Groupings

Parent Group: Surgical Procedures on the Cardiovascular System

- Surgical Procedures on the Heart and Pericardium
 - Surgical Procedures on the Pericardium
 - Excision Procedures of Cardiac Tumor
 - Transmyocardial Revascularization Procedures
 - Pacemaker or Pacing Cardioverter-Defibrillator Procedures
 - Electrophysiologic Operative Procedures on the Heart and Pericardium
 - Introduction or Removal of Patient-activated Cardiac Event Recorder
 - Surgical Procedures on the Heart (Including Valves) and Great Vessels
 - Surgical Procedures on Cardiac Valves
 - * Surgical Procedures on the Aortic Valve (22 procedures)
 - * Surgical Procedures on the Mitral Valve (8 procedures)
 - * Surgical Procedures on the Tricuspid Valve (5 procedures)
 - * Surgical Procedures on the Pulmonary Valve (7 procedures)
 - Other Cardiac Valvular Procedures
 - Coronary Artery Anomaly Procedures
 - Endoscopy Procedures on the Heart and Pericardium
 - Venous Grafting Only for Coronary Artery Bypass
 - Combined Arterial-Venous Grafting for Coronary Bypass
 - Arterial Grafting for Coronary Artery Bypass
 - Coronary Endarterectomy Procedures
 - Repair Procedures for Single Ventricle and Other Complex Cardiac Anomalies
 - Repair Procedures for Septal Defect
 - Repair Procedures for the Sinus of Valsalva
 - Repair Procedures for Venous Anomalies
 - Shunting Procedures on the Heart and Pericardium

- Repair Procedures for Transposition of the Great Vessels
 - Repair Procedures for Truncus Arteriosus
 - Repair Procedures for Aortic Anomalies
 - Repair Procedures for Thoracic Aortic Aneurysm
 - Endovascular Repair Procedures of the Descending Thoracic Aorta
 - Surgical Procedures on the Pulmonary Artery
 - Heart/Lung Transplantation Procedures
 - Extracorporeal Membrane Oxygenation or Extracorporeal Life Support Services and Procedures
 - Cardiac Assist Procedures
 - Other Cardiac Surgery Procedures
- Surgical Procedures on the Arteries and Veins

A.2 Specialties included in matched sample

Cardiac surgery/cardiology, emergency medicine, general surgery, interventional cardiology, interventional radiology, nephrology, neurosurgery, orthopedic surgery, otolaryngology, plastic and reconstructive surgery, pulmonary disease, surgical oncology, urology, vascular surgery.

Appendix B Appendix to Theoretical Framework

B.1 Calculation of Work Hours Adjustment

On average, female physicians work fewer hours than male physicians. As such, women see approximately 0.67 patients for every patient a man sees (Medscape Physician Compensation Report, 2016). In the Medicare data, female surgeons receive 10% of referrals, meaning that they receive 0.1 referrals when working $\frac{2}{3}$ of a “man day” (number of hours men work on average). If we increase a woman’s work hours to be equivalent to a man’s, women should receive 0.05 additional referrals (since they receive 0.1 referrals per $\frac{2}{3}$ day worked, they receive 0.05 referrals per $\frac{1}{3}$ day worked).

Women should thus receive 15% of total referrals.

B.2 Proofs

Proof of Proposition 1

Assume that statements 1 and 2 in Proposition 1 hold and that $\frac{\mathbb{P}(s_G|w)}{\mathbb{P}(s_G|m)} < 1$. Statement 2 says that

$$\bar{a}_w - \bar{a}_m > \mathbb{E}(a|s_G, w) - \mathbb{E}(a|s_G, m) \quad (18)$$

$$\bar{a}_w - \bar{a}_m > \mathbb{E}(a|s_B, w) - \mathbb{E}(a|s_B, m) \quad (19)$$

Multiplying equation 18 by $\mathbb{P}(s_G|w)$ and equation 19 by $\mathbb{P}(s_B|w)$ and adding the two inequalities together gives

$$\begin{aligned} \mathbb{P}(s_G|w)(\bar{a}_w - \bar{a}_m) + \mathbb{P}(s_B|w)(\bar{a}_w - \bar{a}_m) > \\ \mathbb{P}(s_G|w)(\mathbb{E}(a|s_G, w) - \mathbb{E}(a|s_G, m)) + \mathbb{P}(s_B|w)(\mathbb{E}(a|s_B, w) - \mathbb{E}(a|s_B, m)) \end{aligned}$$

Since $\mathbb{P}(s_G|w) + \mathbb{P}(s_B|w) = 1$, we have

$$\begin{aligned} \bar{a}_w - \bar{a}_m &> \mathbb{P}(s_G|w)(\mathbb{E}(a|s_G, w) - \mathbb{E}(a|s_G, m)) + \mathbb{P}(s_B|w)(\mathbb{E}(a|s_B, w) - \mathbb{E}(a|s_B, m)) \\ \bar{a}_w - \bar{a}_m &> \mathbb{P}(s_G|w)\mathbb{E}(a|s_G, w) + \mathbb{P}(s_B|w)\mathbb{E}(a|s_B, w) - \mathbb{P}(s_G|w)\mathbb{E}(a|s_G, m) - \mathbb{P}(s_B|w)\mathbb{E}(a|s_B, m) \\ \bar{a}_w - \bar{a}_m &> \bar{a}_w - \mathbb{P}(s_G|w)\mathbb{E}(a|s_G, m) - \mathbb{P}(s_B|w)\mathbb{E}(a|s_B, m) \\ \bar{a}_m &< \mathbb{P}(s_G|w)\mathbb{E}(a|s_G, m) + \mathbb{P}(s_B|w)\mathbb{E}(a|s_B, m) \end{aligned} \quad (20)$$

where the law of iterated expectations is used in lines 3-4. Substituting in for \bar{a}_m , using the fact that $\mathbb{P}(s_B|m) = 1 - \mathbb{P}(s_G|m)$, and rearranging gives

$$\begin{aligned} \mathbb{P}(s_G|m)\mathbb{E}(a|s_G, m) + \mathbb{P}(s_B|m)\mathbb{E}(a|s_B, m) &< \mathbb{P}(s_G|w)\mathbb{E}(a|s_G, m) + \mathbb{P}(s_B|w)\mathbb{E}(a|s_B, m) \\ \mathbb{P}(s_G|m)\mathbb{E}(a|s_G, m) + (1 - \mathbb{P}(s_G|m))\mathbb{E}(a|s_B, m) &< \mathbb{P}(s_G|w)\mathbb{E}(a|s_G, m) + (1 - \mathbb{P}(s_G|w))\mathbb{E}(a|s_B, m) \\ \mathbb{P}(s_G|m) \cdot (\mathbb{E}(a|s_G, m) - \mathbb{E}(a|s_B, m)) &< \mathbb{P}(s_G|w) \cdot (\mathbb{E}(a|s_G, m) - \mathbb{E}(a|s_B, m)) \\ (\mathbb{P}(s_G|m) - \mathbb{P}(s_G|w)) \cdot (\mathbb{E}(a|s_G, m) - \mathbb{E}(a|s_B, m)) &< 0 \end{aligned} \quad (21)$$

For equation 21 to be negative, one of the terms must be negative and the other positive. Note that $(\mathbb{E}(a|s_G, m) - \mathbb{E}(a|s_B, m))$ must be positive by the definition of a good and bad event, so it must be that $\mathbb{P}(s_G|m) < \mathbb{P}(s_G|w)$ which violates our initial assumption that $\frac{\mathbb{P}(s_G|w)}{\mathbb{P}(s_G|m)} < 1$.

Proof of Proposition 2

Assume without loss of generality that the variance over priors is larger for female surgeons than male surgeons: $\sigma_w^2 > \sigma_m^2$. The law of total variance states that the prior variance is the sum of the variance of the posterior and the (expected) posterior variance:

$$\sigma_g^2 = \text{Var}(\mathbb{E}[a|s, g]) + \mathbb{E}[\sigma_g^2|s]$$

Taking the difference in the variance of the prior between men and women gives

$$\sigma_w^2 - \sigma_m^2 = \mathbb{E}[\sigma_w|s] - \mathbb{E}[\sigma_m|s] + \text{Var}(\mathbb{E}[a|s, w]) - \text{Var}(\mathbb{E}[a|s, m]) \quad (22)$$

We have assumed that $\sigma_w^2 - \sigma_m^2 > \mathbb{E}[\sigma_w|s] - \mathbb{E}[\sigma_m|s]$. Substituting 22 into this inequality, we get

$$\begin{aligned} \mathbb{E}[\sigma_w|s] - \mathbb{E}[\sigma_m|s] + \text{Var}(\mathbb{E}[a|s, w]) - \text{Var}(\mathbb{E}[a|s, m]) \\ > \mathbb{E}[\sigma_w|s] - \mathbb{E}[\sigma_m|s] \\ \text{Var}(\mathbb{E}[a|s, w]) - \text{Var}(\mathbb{E}[a|s, m]) > 0 \end{aligned} \quad (23)$$

Equation 23 shows that the spread over beliefs about a female surgeon after a signal is larger than the spread in beliefs about a male surgeon after the equivalent signal, meaning that the physician must have updated more about the female surgeon after both good and bad signals.

Appendix C Additional Tables

TABLE A1: BALANCE FOR PLACEBO MATCHED SAMPLES

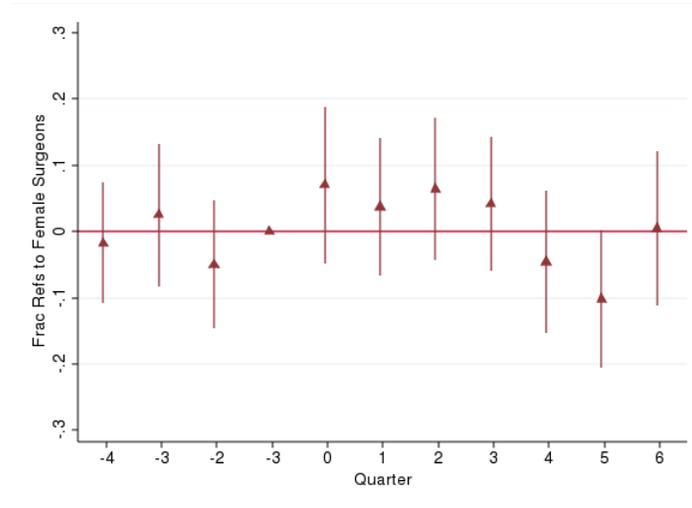
<i>Panel A: Male Surgeons</i>	No Death		Death		p-value
	Mean	SD	Mean	SD	
Patient Refs from Physician	25.6	33.4	25.4	34.2	0.654
Total Patient Refs	69.3	106.0	69.0	105.9	0.899
Patient Age	77.8	9.0	77.9	9.1	0.451
Patient Minority (%)	6.8	25.2	6.8	25.2	0.999
Patient Female (%)	59.0	49.2	59.0	49.2	0.999
Patient Risk (%)	0.85	0.34	0.85	0.35	0.156
Risk All Past Ptnts	0.010	0.007	0.010	0.006	0.194
Experience (Yrs)	24.9	10.4	24.9	10.3	0.997
Available Surgeons	48.7	29.7	49.0	29.6	0.123
Observations	15,012		15,012		

<i>Panel B: Female Surgeons</i>	No Death		Death		p-value
	Mean	SD	Mean	SD	
Patient Refs from Physician	25.4	44.4	25.1	44.2	0.664
Total Patients Refs	72.3	110.0	72.0	107.7	0.821
Patient Age	79.7	9.8	79.8	9.8	0.317
Patient Minority (%)	5.7	23.2	5.7	23.2	0.999
Patient Female (%)	61.2	48.7	61.2	48.7	0.999
Patient Risk (%)	0.010	0.004	0.010	0.005	0.162
Risk All Past Ptnts	0.009	0.004	0.009	0.004	0.284
Experience (Yrs)	22.2	8.3	22.2	8.5	0.640
Available Surgeons	48.8	29.7	48.7	29.4	0.254
Observations	4,658		4,658		

Notes: This balance table shows summary statistics for the matched sample of surgeons who did and did not experience a patient death. The surgeons who did not experience a death are called “placebos” in the paper. To create each sample, I take the set of male or female surgeons who ever experienced a bad event and match the male surgeons to other men who did not experience a death, and match the female surgeons to other women who did not experience a death. The sample is matched exactly on patient gender and minority status as well as surgeon specialty and procedure (not shown above). The sample is matched coarsely on all other variables. All variables are defined as in Table 2.

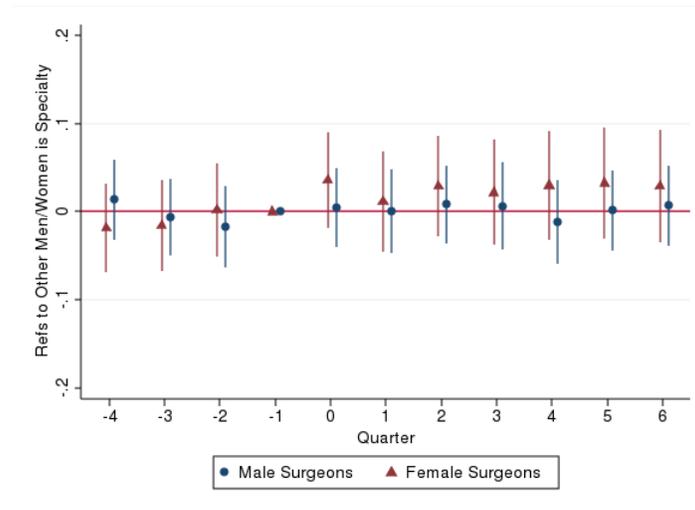
Appendix D Additional Figures

FIGURE A1: SPILLOVERS TO FEMALE SURGEONS, >10 PRIOR REFS



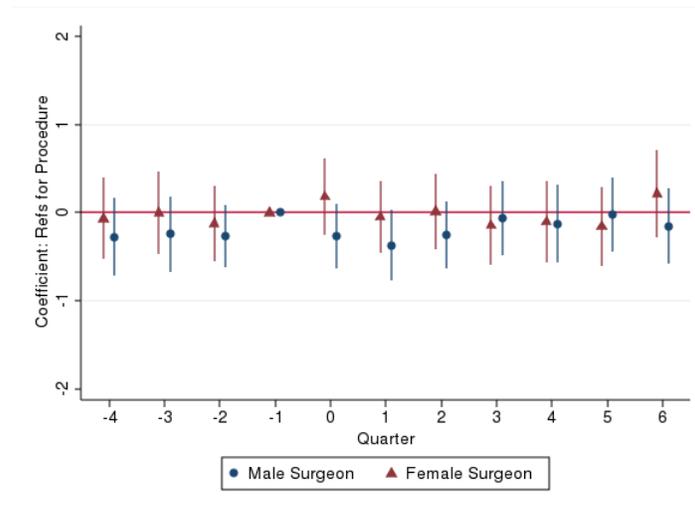
Notes: This figure shows how physicians change their behaviour toward other female surgeons after a patient death. In $k = 0$, a patient that physician j sent to surgeon i dies. The outcome variable is the fraction of referrals going to female surgeon whom physician j had referred at least 10 patients to prior to the death. The estimation is done on the sample of female surgeons who experience a patient death. The coefficients are plotted relative to the fraction of referrals that physician j was sending to these surgeons in quarter $k = -1$. I control for the fraction of available surgeon who are male or female and also include physician-surgeon match fixed effects. Standard errors are clustered at the physician-surgeon match level.

FIGURE A2: SPILLOVERS TO OTHER SURGEONS AFTER GOOD OUTCOME



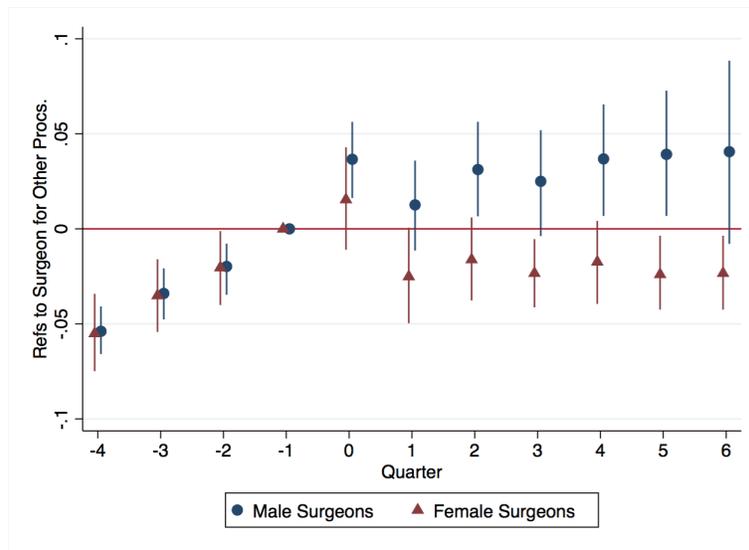
Notes: This figure shows how physicians change their behaviour toward other surgeons after an unexpectedly good outcome. In $k = 0$, a risky patient that physician j sent to surgeon i is not rehospitalized within 30 days of surgery. The outcome variable is the number of referrals going to female (Figure A) or male (Figure B) surgeons whom physician j hasn't referred to before. The coefficients are plotted relative to the number of referrals that physician j was sending to surgeons she had not previously referred to in quarter $k = -1$. Physician-surgeon match fixed effects are included in the regression and standard errors are clustered at the physician-surgeon match level.

FIGURE A3: REFERRALS FOR PROCEDURE



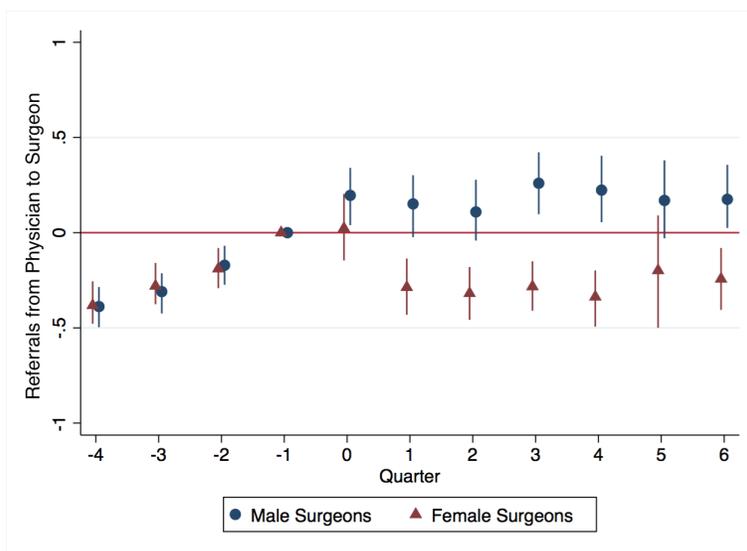
Notes: This figure plots the coefficients from a regression estimating how many patient referrals a physician gives for the procedure that a patient dies from in quarter $k = 0$. The outcome variable is the quarterly number of referrals for the particular procedure that physician sends to surgeons other than the performing surgeon. The estimation is done on the sample of matched male and female surgeons who experience a patient death. All coefficients are plotted relative to the number of procedural referrals the physician was sending in $k = -1$, which is normalized to zero. Surgeon-physician match fixed effects are included and standard errors are clustered at the physician-surgeon match level.

FIGURE A4: REFERRALS FOR OTHER PROCS. TO PERFORMING SURGEON



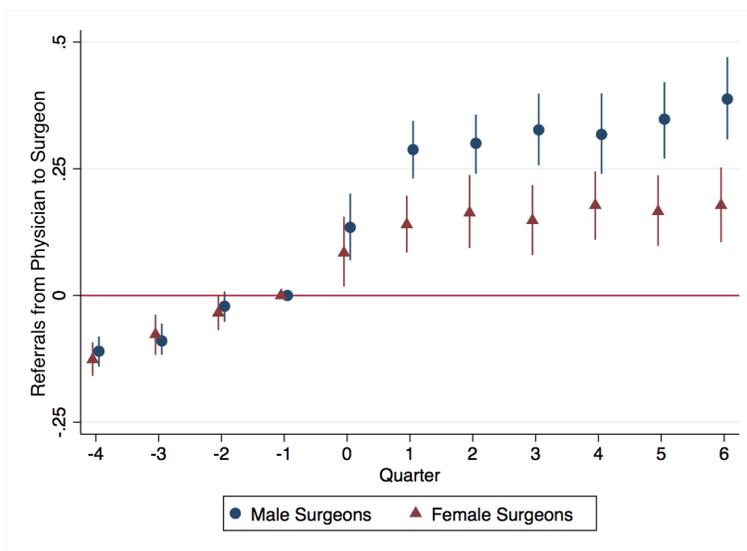
Notes: This figure shows the number of patients that physician j refers to surgeon i for a procedure other than the procedure that the patient who died was referred for. In $k = 0$, a patient that j referred to i for procedure n dies. The outcome variable is the number of referrals that j refers to i in each quarter for any procedure other than n . The coefficients are plotted relative to $k = -1$ and 95% confidence intervals are also shown. The coefficients for $k = 1$ through $k = 6$ are jointly significantly different from zero at the 10% level for both male and female surgeons. The sample used is the sample of matched surgeons who experienced a patient death. Surgeon-physician match fixed effects are included. Standard errors are clustered at the physician-surgeon match level.

FIGURE A5: DEATHS THAT OCCUR ON DAY OF SURGERY



Notes: This figure plots the quarterly regression coefficients and 95% confidence intervals from estimating equation 3 using the sample of matched male and female surgeons who experience a patient death that occurs within 24 hours of surgery. The coefficients are plotted relative to the number of referrals the surgeon was receiving from the physician in $k = -1$, which are normalized to zero. In $k = -1$, male and female surgeons both received an average of 0.65 referrals from physician j . A patient that physician j referred to surgeon i dies in $k = 0$. The outcome variable is the total number of referrals that physician j sends to surgeon i each quarter. Standard errors are clustered at the physician-surgeon match level.

FIGURE A6: UNEXPECTEDLY GOOD OUTCOMES: TOP 5% RISK LEVEL



Notes: This figure plots the quarterly coefficients and 95% confidence intervals from estimating equation 3, looking at good events. Here, good events are defined using a 5% risk cutoff rather than a 1% risk cutoff. All coefficients are plotted relative to the number of referrals the surgeon was receiving from the physician in $k = -1$, which are normalized to zero. Standard errors are clustered at the physician-surgeon match level.