Gender Mix and Team Performance: Evidence from Obstetrics*

Ambar La Forgia
UC Berkeley & NBER
ambar@berkeley.edu

Manasvini Singh
Carnegie Mellon University
msingh01@cmu.edu

Abstract

We investigate how the gender mix of expert teams affects performance in a high-stakes setting: childbirth. Using data on 2.5 million births, we exploit the quasi-exogenous assignment of patients to two-member obstetrician teams (Lead–Assisting), and find that: (i) female-only teams achieve the best maternal outcomes, whereas male-only teams have the worst; and (ii) female-led mixed-gender teams perform worse than male-led ones. These patterns cannot be explained by patient risk, endogenous team formation, or physician preferences for discretionary practices like C-sections. Instead, gender mix directly affects team decisions and performance, likely through gender norms — a mechanism supported by two findings. First, gender mix affects how closely team decisions reflect member preferences, with female-only teams being especially skilled at this process, possibly due to more collaborative decision-making. Second, gender mix affects team resilience, with female-led mixed gender teams performing especially poorly under challenging conditions (e.g., limited team familiarity), possibly because female leaders invert traditional gender norms.

*We gratefully acknowledge feedback from seminar participants at the London School of Economics, University of Southern California, University of British Columbia, Cornell University (Economics Department), Cornell University (Weill Cornell School of Medicine), Behavioral Economics Design Incentive (BEDI) Workshop at the University of Pittsburgh, West Virginia University, Center for Health Incentives and Behavioral (CHIBE)) at Penn Research Seminar, ASHECon 2025, Berkeley Haas, and thank Dave Chan, Mathijs De Vaan, Adam Leive, and Jacob Zureich for their helpful feedback. We are especially grateful to Solène Delecourt for her contributions to earlier versions of this paper.

1 INTRODUCTION

Teamwork is integral to performance in many professional settings, such as business, law, and medicine. Several factors can affect team performance beyond the mere sum of the individual abilities of team members. One factor influencing team performance is the gender mix of the team, which has been shown to affect the collaborative process of teams within classroom and online laboratories.[1] While these studies have established a rich theoretical foundation for understanding gender dynamics in teams, the dynamics between laypersons (e.g., students or MTurk participants) in artificial environments may not translate to complex "real world" environments where professionals make high-stakes decisions that require teams to communicate and integrate diverse views and expertise under pressure. This paper bridges the gap by studying how the gender mix of expert teams affects team performance in a high-pressure environment with potentially life-or-death consequences: childbirth.

Physicians are highly trained medical experts who often have to work together to provide quality care to patients while balancing competing goals, practice styles, experiences, and resource constraints. These differences can create substantial friction when working in teams (Chan, 2021; Chen, 2021; Silver, 2021), particularly in high-stakes settings such as childbirth. For instance, the unpredictable nature of labor often requires rapid and subjective decision-making, and the ability to perform under the risk of harm to two patients: the mother and the baby. Obstetrics is also one of the most gender-diverse surgical specialties (nearly 60% of practicing obstetrician-gynecologists (OB-GYNs) are women (Boyle, 2021)), creating frequent opportunities for physicians of the same or opposite gender to work together. Research has found that male and female physicians practice medicine differently (Ganguli et al., 2020; Greenwood, Carnahan, & Huang, 2018), which may create synergies that benefit same-gender teams. However, mixed-gender teams may also benefit from diversity in decision-making. Therefore, it remains an empirical question how team gender dynamics shape patient outcomes.

We use inpatient discharge data on the universe of births in Florida from 2006-2018 to study the causal effect of team gender mix on patient outcomes. While most births record a single OB-GYN monitoring the patient and delivering the baby, during 23% of births (570,982 births), a second OB-GYN is documented as participating in the labor and delivery process. Therefore, we construct a "team sample" comprising of deliveries where there is both a "Lead" and an "Assisting" physician, creating four distinct types of teams: two same-gender teams (male-only and female-only) and two mixed-gender teams (female-led and male-led). Our primary outcome of interest is whether a patient (the mother) experiences a severe complication during labor and delivery that "results in significant short- or long-term consequences to a woman's health" (ACOG, 2016b).[2] This is a critical outcome to study because maternal complication rates remain alarmingly high even in the United States (Khalil et al., 2023), and OB-GYNs, as the key providers during most births, can directly influence these maternal outcomes. For example, use of interventions such as cesarean sections (C-sections) can vary widely across physicians and are closely tied to patient outcomes (Corredor-Waldron, Currie, &

[1] Most of the literature in economics and psychology studying gender and team behavior have relied on classroom experiments (Aparicio Fenoll & Zaccagni, 2022; Apesteguia, Azmat, & Iriberri, 2012; Hoogendoorn, Oosterbeek, & van Praag, 2013; Karpowitz et al., 2024), and laboratory or online experiments with small payoffs (Born, Ranehill, & Sandberg, 2022; Gneezy, Niederle, & Rustichini, 2003; Hardt, Mayer, & Rincke, 2024; Ivanova-Stenzel & Kübler, 2005; Kearney et al., 2022; Li et al., 2022; Mendelberg, Karpowitz, & Goedert, 2014; Woolley et al., 2010). While these studies allow for precise development and testing of theory, they are more limited in their generalizeability to real-world settings involving expert professionals. Studies that examine gender and team dynamics in naturalistic settings largely focus on innovation (e.g., team gender mix and novelty and impact of publications (Yang et al., 2022)) and corporate settings (e.g., gender quotas in European companies and performance (Bertrand et al., 2014; Latura & Weeks, 2023; Lu, Naik, & Teo, 2024; Matsa & Miller, 2013)). However, these settings have inherently non-random team formation, which can affect performance independent of team gender mix.

[2] Specifically, we create a binary indicator based on a standard metric for whether a patient ever experiences a severe complication during labor and delivery, such as acute respiratory distress or cardiac failure (Main et al., 2016).

Schnell, 2024; La Forgia, 2023; Singh, 2021). Overall, if the gender mix of the team affects the team's obstetric decisions, there should be discernible differences in maternal complications across teams.

Exploiting the quasi-random assignment of patients to two-member obstetrician teams of varying gender mix, we find that same-gender teams improve performance for female physicians but harm performance for male physicians. We focus on two main results. First, deliveries by female-only teams have the lowest rates of complications (2.46%) of all four types of teams, while those by male-only teams have the highest (2.88%). We calculate that had female-only teams performed all births in the team sample, maternal complication rates would be 8% lower. Second, among mixed-gender teams, we find that female-led teams perform worse than male-led teams. These results are obtained from regressions that control for (i) 23 antepartum patient risk factors, including pre-existing conditions such as hypertension or diabetes that could affect outcomes, (ii) patient race/ethnicity and insurance status, which could influence access to prenatal care and socioeconomic factors affecting health, (iii) physician age, experience and collaborative history, which helps account for skill development and familiarity with other physicians, and (iv) hospital × year and quarter of birth fixed effects to account for unobserved factors that could vary across hospitals or time, such as changes in hospital policies, staffing, or seasonal birth trends.

We demonstrate that our results are robust to alternative specifications, samples, and outcome variables. First, interpretations are qualitatively similar when using a regression with only year and hospital fixed effects, to a saturated model with all aforementioned controls, as well as patient zip code fixed effects and quarter × year × hospital × day-of-week × hour-of-day fixed effects. Second, we find similar results whether we restrict our sample to only C-sections or only vaginal births (though complication rates are over 3 times higher for C-sections), as well as to samples excluding planned or prior C-sections, potential residents, teaching hospitals, maternal fetal medicine specialists, and hospitals without at least one of each type of team. Third, we show results are robust to using more expansive and more restrictive definitions of maternal complications. Lastly, as a more precise test of mixed-gender team performance, we include a fixed effect for the physician team and find that for the *same* two physicians, maternal complications are significantly higher when the female is the Lead than when the male is the Lead.

To interpret these findings as causal, patient assignment to the *gender mix* of the team should be as-random, conditional on covariates. (We do not require the stricter assumption of random assignment of a patient to a *team* for identification.) Conceptually, we find that this assumption is supported by informal interviews with OB-GYNs as well as documentation provided by hospitals, professional organizations, and popular media. That is, the randomness in birth timing, patient circumstances during childbirth, and physician schedules create limited opportunities for either the patient or the physician to select into a team based on the genders of the physicians. Empirically, in support of the identifying assumption, we show that the four types of teams treat patient populations of similar risk, and that teams do not appear to be formed endogenously according to physician gender, familiarity, skill, or patient risk. For example, we show that there are no differences between male and female Leads in the gender of the Assisting they are paired with, regardless of patient risk. Together, these results minimize concerns of patient or physician selection and support a causal interpretation of the effect of team gender mix on maternal complications.

Heterogeneity in outcomes by patient risk, race, and insurance status provide further support of our findings. For example, a patient's observed medical risk factors could influence both their outcomes and assignment to a certain team. However, in subsample analyses we find similar patterns for both

low and high-risk patients, with female-only teams achieving lower complication rates regardless of patient risk. Next, we consider heterogeneity by patient race, given the extensive research documenting racial disparities in maternal care (Alsan & Yearby, 2024; Corredor-Waldron, Currie, & Schnell, 2024; Hill et al., 2024), and find that female-only teams are unique in achieving both the lowest overall complication rates and no disparity between Black and non-Black patients. Male-only teams exhibit the opposite pattern, with the highest complication rates and the largest racial disparities. Lastly, we show that our findings are similar for both privately and publicly-insured patients, suggesting financial incentives are not driving observed differences by team gender mix.

Next, we examine the potential mechanisms underlying this causal relationship. We begin by asking: *how* does gender mix affect team performance? To do so, we focus on the clinical decision to perform a C-section, as it directly influences maternal complications (both in our data and prior literature (Costa-Ramón et al., 2018; Yu et al., 2023)). Moreover, its discretionary nature results in considerable variation in baseline use across individual physicians. This allows us to examine how gender mix affects a team's decision-making process, especially when team members may have differing preferences for C-sections. We distinguish between two pathways through which gender mix can influence team outcomes, as each pathway has different policy implications: the "Individual Preference" effect – where differences in team decisions simply reflect differences in C-section preferences between male and female physicians[3] – and the "Gender Mix" effect – where team decisions are directly and causally impacted by the genders of the two physicians on the team, holding their preferences fixed.

We proxy for physicians' individual preferences for C-sections by isolating their C-section fixed effects from the sample of deliveries for which they are the only physician on the record. These individual physician preferences strongly predict both team C-section decisions and maternal complication rates. We also find that male physicians exhibit stronger preferences for C-sections than female physicians. However, even holding the preferences of the Lead and Assisting physician constant, we still find stark differences in team decisions and performance across teams of varying gender mix. This result provides strong support for the "Gender Mix" effect, which suggests that for the same team, changing nothing but the gender of the team members could alter team decisions and performance.

Next we ask: *Why* would a team's gender mix directly and causally affect their decisions and performance? We hypothesize that gender mix introduces social norms around gender (i.e., gender norms), which can influence how effectively teams communicate and integrate diverse viewpoints, how collaboratively versus hierarchically decisions are made, or how they navigate challenges. While we cannot measure these norms directly, two data patterns support our hypothesis.

First, gender mix affects how individual physician preferences are incorporated into team decisions. We find that unlike mixed-gender teams, same-gender teams make C-section decisions that closely reflect the average preferences of their members, potentially reflecting a more collaborative decision-making process. We also find that all teams *except* female-only teams default towards performing more C-sections than the individual physicians would prefer to perform by themselves. Teams with a male physician may default to C-sections because they are often viewed as the "path of least resistance" and a more decisive, action-oriented choice – and gender norms in team settings may amplify a man's risk-tolerant behaviors while simultaneously undervaluing women's input (Coffman, Flikkema, & Shurchkov, 2021; Lamiraud & Vranceanu, 2018; Thomas-Hunt & Phillips, 2004). Female-

[3] A physician's C-section preference can be viewed as their inherent practice style or personal threshold for choosing a C-section, independent of external factors. If there are gender-based differences in preferences, then preferences – not gender itself – may explain variation in decisions and outcomes across team types. For instance, if male physicians generally prefer C-sections and female physicians prefer vaginal deliveries, then we should expect male-only teams to have higher C-section rates than female-only teams on average, with mixed-gender teams falling between the two.

only teams, less constrained by such norms, appear to uniquely resist defaulting to more C-sections, which may explain their superior performance across all four types of teams.

Second, gender mix affects team resilience to challenging conditions that may make collaboration more difficult. We study four such challenges: (i) conflicting preferences for C-section among team members; (ii) the Lead physician being notably younger than the Assisting physician; (iii) a limited history of prior collaboration between team members; and (iv) delivering during periods of high hospital strain. We find that female-led mixed-gender teams are the least resilient: they are the only team that consistently performs worse under *each* of these challenging conditions, perhaps because having female leaders overseeing male subordinates inverts traditional gender norms on leadership. This inversion may introduce additional friction into team interactions, placing these teams at an inherent disadvantage from the outset (Beaman et al., 2009; Eagly, Makhijani, & Klonsky, 1992).

This paper makes two primary contributions to the literature on gender dynamics in teams. First, previous studies typically either rely on experiments in artificial environments (e.g., labs or classrooms), establishing causality but lacking real-world applicability, or use observational data, which provide realistic contexts but cannot credibly establish causality due to endogenous team formation (see Footnote 1 for citations). We fill this gap by providing quasi-experimental evidence from a high-stakes, real-world context with unique features - such as informal hierarchy – that add important nuance to the study of gender dynamics in teams. Second, our study leverages a unique innovation unavailable to prior research: detailed data on individuals making decisions by themselves *and* as part of teams. This enables us to precisely track how gender mix moderates the transition from individual decision-making to team-based decision-making, allowing us to offer a unifying explanation for the mixed literature on this topic: whether same- or mixed-gender teams perform better depends critically on the social norms shaping the decision-making environment.

Finally, our findings generate practical insights with relevance beyond obstetrics. Other high-stakes settings, from lawyers working to reverse a wrongful conviction to pilots flying through a storm, also rely on expert professionals collaborating to achieve a desired outcome. Like medicine, these traditionally male-dominated professions are becoming more diverse, potentially introducing gender-based frictions that could affect performance. In our setting, we show that performance is not just a product of each team member's individual preferences but a direct result of how male and female team members interact. While organizations can shift individual preferences through training or incentives, social norms around gender are inherently more challenging to address, requiring deliberate processes to reduce interpersonal friction or support adversely affected groups, such as female-led mixed-gender teams. Sorting teams based on gender may also achieve desired outcomes, but imposes ethical and psychological costs that limit feasibility. Ultimately, our research demonstrates the critical impact of gender mix on team performance, underscoring the need for more research on gender and team dynamics in an increasingly diverse workforce.

2 METHODS

2.1 Data and Setting

We use de-identified patient-level data from hospital discharge records from the Florida Agency for Healthcare Administration. These data, which have been used extensively in healthcare research, include all births delivered in Florida hospitals between 2006 and 2018 (Greenwood, Carnahan, & Huang, 2018; La Forgia, 2023; Zureich & Singh, 2024). Each observation is a birth record, which

includes information on the mother's age, race, insurance status, zip code, procedure, and diagnosis codes, as well as information on the hospital and the physician(s) responsible for the patient's care. Physician characteristics, including age, gender, and specialty, were obtained from three different sources: (i) publicly available Florida Licensure data (data that physicians are required to provide to practice in Florida), (ii) Medicare's Physician Compare National Downloadable File (data that physicians are required to provide to participate in Medicare), and (iii) from the proprietary SK&A Physician Survey (a comprehensive survey of office-based physicians, rebranded as IQVIA OneKey). In Appendix A, we provide more details on all datasets.

Table 1 presents summary statistics for our sample, and Appendix Figure A.1 depicts our sample restrictions.[4] The final sample includes 2,507,736 births delivered by OB-GYNs, of which 1,936,754 are delivered by a single OB-GYN (i.e., the "solo" sample) and 570,982 births have a team of two OB-GYNs involved in the labor and delivery process (i.e., the "team" sample). The team sample is our main analytic sample, covering the 23% of births in Florida where two OB-GYNs were recorded as having responsibility for the patient during labor and delivery. However, we use the solo sample to support our identifying assumption (e.g., estimate physician skill) and explore mechanisms (e.g., estimate physician preferences) as all physicians who deliver as a team also deliver by themselves.

Using these inpatient discharge records, we can identify the "practitioner who had primary responsibility for the patient's medical care and treatment or who certified as to the medical necessity of the services rendered," which we refer to as the "Lead" physician, and the "practitioner who had primary responsibility for the principal procedure performed", which we refer to as the "Assisting" physician (The Office of Data Dissemination and Transparency, n.d.). When the Lead and Assisting physician have the same medical license, this is a "solo" birth, and when they are different, this is a "team" birth. The principal procedure could be delivering the baby or other interventions during labor and delivery, which the Assisting physician should not, in theory, be able to perform without the Lead physician's approval. Therefore, both physicians have responsibility for the patient, but the Lead physician is ultimately the one liable for the patient's care under state law.[5] See Appendix B for more details on physician roles and responsibilities.

We will use the term "team types" when referring to the four combinations of Lead (L) - Assisting (A) pairs: two same-gender teams (M_L–M_A and F_L–F_A) and two mixed-gender teams (M_L–F_A and F_L–M_A). In the team sample, there are 1,010 female OB-GYNs and 1,034 male OB-GYNs (2,044 total), with female OB-GYNs recorded as the Lead in 42% of births (Appendix Table A.1). Each OB-GYN occupies both roles, meaning that during some births the OB-GYN is the Lead, and during other births they are the Assisting.

Two OB-GYNs may participate in the same birth for a variety of reasons. A common scenario is that the Lead physician requires assistance during a more complicated delivery. Another scenario is that the Lead physician monitored the patient up until delivery, but the Assisting physician performed the delivery because of a scheduling change or the Lead attending another birth. The Assisting

[4]The main sample restrictions are as follows: starting with 2,731,325 births with a physician on the record, we drop hospitals delivering fewer than 100 births per year (692 births), all births that also include a nurse or midwife on the discharge record (110,075 births), births where a physician on the discharge record is not an OB-GYN (22,361 births), births where there is a unique third physician on the discharge record (58,112 births), OB-GYNs delivering fewer than 36 births a year on average, which represents the bottom 1% in yearly volume (30,858 births), OB-GYNs who are not observed in both a "Lead" and "Assisting" role (1,408 births), and births where patients have outlier or wrongly imputed ages (83 births).

[5]In inpatient discharge data, the Lead physician is the attending physician, and they ultimately bear the legal responsibility of medical decisions, document all care provided for the patient, and are viewed as the leader of the team (Houchens et al., 2020; Kevin, 2020). For example, according to a Florida statute specific to childbirth, "any birth other than a normal birth frequently leads to a claim against the attending physician," highlighting the legal recognition of the attending physician's central role in patient care and associated liability (Florida Legislature, 2023).

physician may also have been the patient's primary OB-GYN and could not arrive at the hospital to admit the patient, but arrived in time to deliver the patient (our data does not allow us to identify the patient's primary OB-GYN, but obstetric medical practices with multiple physicians often have a call schedule where they deliver the patients in labor at that time, not necessarily the patient for whom they provided prenatal care). The key point is that in all scenarios, the two OB-GYNs must actively communicate and coordinate face-to-face during the labor and delivery process, even if both are not present in the delivery room at the exact time of the birth. For instance, during a patient hand-off, physicians are expected to communicate "up-to-date information regarding patient care, treatment, and service, condition, and any recent or anticipated changes," and this communication "should be interactive to allow for discussion between those who give and receive patient information" (ACOG, 2012). Therefore, the quality of care a patient receives when treated by a team is the result of the joint performance of the physician team during the birth encounter.[6]

2.2 Outcome Variables

Our main outcome variable is whether a patient experiences a maternal complication during labor and delivery, identified using patient diagnosis and procedure codes from the International Classification of Diseases, Ninth and Tenth Revision (ICD-9 and ICD-10). Specifically, we follow the standardized metric "severe maternal morbidity" to create an indicator equal to one if a patient ever experiences *any* of 25 serious and acute conditions *during* labor and delivery.[7] This indicator for maternal complications captures "near-misses" caused by life-threatening conditions that can occur during labor and delivery and so does not include complications before birth. These events are considered to have a "high rate of preventability" (ACOG, 2016b) and thus may reflect poor decision-making by the team. Adapting the approach of (Snowden et al., 2021), we show robustness to using both a more restrictive and more expansive measure of maternal complications (Bateman et al., 2013; CDC, 2024) as well as to limiting the sample to before the change to ICD-10 codes.

2.3 Identifying Assumption

The goal of this paper is to estimate the causal effect of a team's gender mix on maternal complication rates. A causal interpretation requires the following identifying assumption: the assignment of patients to the gender mix of the team is quasi-exogenous. This assumption does not require that the assignment of patients to physician *teams* be as-random (which is a more stringent assumption), only that assignment to *team type* (i.e., gender mix) be random. In other words, patients can be assigned to physician teams endogenously, but the observed or unobserved factors that are driving this assignment should be uncorrelated with the gender mix of the team.

We provide both conceptual and empirical support for this assumption. Conceptually, the timing of labor onset is random, obstetric units rely on rotating call schedules, and medical emergencies

[6] A similar perspective is taken by Goradia and Chandrasekaran (2024), who use the same data to study cardiology patient outcomes. They describe two-physician teams as having a "dual decision-making approach," compared to the "single decision-making approach" of solo physicians. However, a key difference is that their physicians include different specialists working together with clearly allocated tasks, whereas we study two equally trained OB-GYNs engaged in truly joint decision-making.

[7] This definition was developed by the Centers for Disease Control and Prevention (CDC) using ICD-9 codes before the transition to ICD-10 codes in October 2015 (Hirai et al., 2022; Main et al., 2016). We use this metric because most births in our sample occurred before October 2015. The 25 conditions include: Acute myocardial infarction, acute renal failure, adult respiratory distress syndrome, amniotic fluid embolism, aneurysm, cardiac arrest/ventricular fibrillation, disseminated intravascular coagulation, eclampsia, heart failure during procedure or surgery, internal injuries of thorax, abdomen, and pelvis, intracranial injuries, puerperal cerebrovascular disorders, pulmonary edema, severe anesthesia complications, sepsis, shock, sickle cell anemia with crisis, thrombotic embolism, blood transfusion, cardio monitoring, conversion of cardiac rhythm, hysterectomy, operations on heart and pericardium, temporary tracheostomy, ventilation. See Appendix C for more details.

can reshuffle pre-scheduled procedures (such as C-section and inductions) with little notice, all of which sharply limit patient choice. This reality is frequently acknowledged by professional clinical organizations: ACOG states that "members of a department, practice, or call group often take turns covering the labor ward ... on a rotating schedule that includes day, night, and weekend shifts" (ACOG, 2016a). Hospital groups communicate similar points: "Doctors, midwives, and nurses work on rotating shifts, and like you, most of us have our own families to care for outside of our careers ... we understand the prospect of not having your regular provider at childbirth might be stressful, so we've put together a list of tips to help [you] prepare ..." (UT Southwestern Medical Center, 2021). Additionally, popular media sources reinforce this message. For example, Cosmopolitan magazine notes in an article titled "20 Things No One Tells You About Giving Birth" that "The person who delivers your baby might not be your doctor" (Ruiz, 2014).

Additionally, having two OB-GYNs involved in a single birth is the exception and not the rule. Therefore, it is unlikely that a patient can exert much selection pressure on the *team* they end up with, because it is difficult to know in advance (for certain) whether there will be one or two physicians overseeing their delivery. Even during scheduled C-sections, a patient's procedure may be delayed to accommodate more urgent deliveries on the same day, which may shuffle the delivery team. To summarize, even if patients know their primary physician's identity, there is no guarantee that physician will deliver their baby. It is even less predictable whether a second physician will be involved – or what their gender might be – since this depends entirely on who happens to be available at delivery. Thus, patients have very limited scope to select the assisting physician's gender.

Similar pressures limit an OB-GYN's ability to know or select the second physician on the team. We had informal conversations with four OB-GYNs working in both private hospitals and academic medical centers to further understand team formation. The following quote was particularly illustrative: "The creation of a team is more often than not not premeditated on labor and delivery as it depends on whoever is assigned to call and back-up call. Teams in emergent situations are often quite random based on prior call assignment. Who the next oncoming OB will be for the following shift is not determined by the provider on call; that is predominantly with previously assigned scheduling assignments."

However, the evidence described thus far is anecdotal. Later, in Section 3.3, we present empirical analyses supporting our identifying assumption. Specifically, we show that patient risk factors for maternal complications do not systematically vary across the four team types, nor do teams appear to form endogenously based on physician gender, patient risk, familiarity, or skill. We intentionally first present our empirical strategy, main results, and robustness checks to clearly establish our baseline findings. We then provide the empirical evidence supporting our identifying assumption by ruling out alternative explanations, thereby strengthening our ability to interpret our main results as identified.

2.4 Empirical Strategy

We examine whether maternal complications differ based on the gender mix of physician teams responsible for labor and delivery. Using births where two physicians appear on the discharge record, we estimate the following equation:

$$\begin{aligned} y_{ijkht} = \beta_1 \text{lead_fem}_j + \beta_2 \text{samegender_assist}_k + \beta_3 (\text{lead_fem}_j \times \text{samegender_assist}_k) \\ + \gamma X_i + \delta Z_j + \phi T_k + \alpha_t + \alpha_{hy} + \varepsilon_{ijkht} \end{aligned} \tag{1}$$

where y is the observed outcome for patient i under the care of Lead physician j and Assisting

physician k in hospital h and quarter-year t. The key coefficient of interest, β_3 captures whether the effect of having a same-gender team on maternal complications depends on whether the Lead physician is male or female. A negative interaction term tells us that going from a mixed-gender to a same-gender team reduces maternal complications relatively more for patients of female physicians than for male physicians. See Appendix D.1 for additional details and interpretations of the other coefficients. In addition to showing the regression output, we also graphically plot the fitted values of our outcomes (i.e., $\widehat{MC_i}$ for maternal complications from Equation 1) for easier inspection of estimates across the four team types.

In this specification, X_i is a vector of 23 patient risk factors that are predictive of maternal complications[8], as well as controls for the patient's race/ethnicity (Black, Hispanic/Latina, White, or other Race) and insurance status (private insurance, Medicaid, or other insurance). We include quintiles of physician age and quintiles of a physician's cumulative births up to the previous quarter for both the Lead physician (Z_j) and Assisting physician (T_k), as these likely also affect patient outcomes. We also include a measure of previous collaboration by including quintiles of the cumulative births up to the previous quarter performed by a physician team. Lastly, we include fixed effects for the number of diagnosis codes recorded in the patient's medical record. This accounts for variations in coding behavior by physicians, which could influence our ability to accurately assess the patient's clinical risk. We present results with and without controls to assess sensitivity to differences in the patient population treated by different team types.

To account for any systematic time-specific factors that could influence birth outcomes, we include quarter-year fixed effects (α_t) as data is released at the quarter-level. We also include hospital-year fixed effects (α_{hy}), which would account for any year-to-year changes in a hospital's staffing or labor and delivery policies that could influence birth outcomes. We use robust standard errors clustered at the Lead physician level. Our main estimator is a linear probability model, which has been used extensively by researchers to study adverse patient events because the estimates are easier to interpret, particularly for interaction terms (Ai & Norton, 2003; Greenwood, Carnahan, & Huang, 2018; Tsugawa et al., 2017). However, we also show the robustness of our results using a logit model.

3 RESULTS

Table 2 summarizes all the main analyses, robustness checks, and exploration of mechanisms. This table provides an easy way to compare and reference all the core findings discussed in the paper.

3.1 Main results

Our central finding is that being in a same-gender team lowers maternal complication rates for female physicians, but increases them for male physicians (interaction coefficient $\beta_3 = -.0047$, p-val < 0.001). As seen in Figure I Subfigure (a), maternal complication rates across team types exhibit the following

[8] The full list includes: Patient age (divided into quintiles of age), asthma, anemia, polyhydramnios or oligohydramnios, maternal physical abnormalities (includes thyroid abnormality, bone or joint disorder, or abnormality of organs and soft tissues of pelvis), blood disorders (includes antepartum hemorrhage, abruptio placentae, placenta previa, uterine rupture, coagulation defects complicating pregnancy and spotting complicating pregnancy), uterine size issue (includes uterine size date discrepancy and cervical shortening), infant size issue (i.e., fetal growth issue affecting the management of the mother), obesity, diabetes or abnormal glucose tolerance, substance abuse or smoking, infectious and parasitic conditions, maternal congenital and other heart disease, known or suspected fetal abnormalities affecting the management of the mother, hypertension complicating pregnancy (includes pre-eclampsia), isoimmunization, premature rupture of membrane or amniotic cavity infection, previous pregnancy, malposition or malpresentation of fetus (includes breech birth), multiple gestation (i.e., twins or above), pre-term birth, previous C-section and an indicator of other conditions and risks (includes excessive weight gain during pregnancy, habitual aborter, renal disease, liver disorders, nerve disorders, and severe urinary tract infection. See Appendix Section C for more details).

rank ordering from highest to lowest: male-only team (M_L–M_A) > female-led mixed-gender team (F_L–M_A) > male-led mixed-gender team (M_L–F_A) > female-only team (F_L–F_A). We estimate that if female-only teams had performed all 570,982 deliveries in our sample (keeping covariates at their mean), maternal complications would be 8% lower. In Subfigures (b) and (c), we show that this pattern in maternal complications by team gender mix is the same in births delivered vaginally *and* births delivered by C-section. However, the complication rate is over 3 times higher for C-sections, which we will explore further in Section 4.

To summarize, among same-gender teams, female-only teams exhibit the best performance: the rate of maternal complications is 2.40% for female-only teams versus 2.79% for male-only teams (p-val on difference < 0.001). However, among mixed-gender teams, female-led teams perform worse than male-led teams. While the difference is not significant in this specification, we later show that it becomes significant when using a more precise test with team fixed effects. The rest of our paper will focus on these two patterns: the difference in maternal complications between the two same-gender teams, and the difference between the two mixed-gender teams.

3.2 Robustness checks

Sparse vs. saturated controls: We present robustness checks for our main estimates in Table 3, where Column 1 provides estimates from the main specification and Columns 2-6 demonstrate the stability across specifications ranging from sparse to saturated. Our key finding – that same-gender teams benefit female physicians more than male physicians – is robust to a specification where the *only* control variables we have are hospital and year fixed effects (Column 2). Results remain robust as we sequentially add in patient-level controls (Column 3), physician-level controls (Column 4), and patient zipcode fixed effects that act as a proxy for geographic-level socioeconomic and demographic characteristics that could influence access to care (Column 5). Lastly, results remain robust to including hospital × year × quarter × day-of-week × admission-hour fixed effects (totaling approximately 86,000 FEs), which control for hospital-specific staffing, resources, or clinical schedules on particular weekdays or at certain times of day, such as planned C-sections sections being less common on weekends or in the evenings (Column 6). Note that admission time is only available starting in 2010, and despite this data loss, we continue to find similar effects as the main specification (Column 1).

The stability of our results across these diverse specifications suggests that omitted variable bias plays a limited role in our setting and supports our identifying assumption of quasi-random assignment of patients to teams of varying gender mix. In other words, if patients are quasi-randomly assigned to team types, then raw differences in outcomes should closely match estimates from models that extensively control for patient and physician characteristics, which is precisely what we find.

Logistic regression: Given the rarity of the maternal complication outcome, we show in Appendix Table A.2 that the interpretation of the results remains the same when using logistic regression.

Varying analytic samples: In Appendix Table A.3 we show that our results remain robust under various sample restrictions designed to strengthen our identification strategy. First, we show results are robust to excluding planned C-sections (coded as C-sections without indications of labor, see Appendix C for details) and patients with prior C-sections, as these cases may involve greater scheduling discretion by physicians or patients (Column 1). Second, we show robustness to excluding births involving maternal-fetal medicine specialists, who are disproportionately male and handle high-risk pregnancies – a scenario where patient assignment to teams may be more endogenous (Column 2). Third, we

show robustness to excluding births involving any physician younger than 33 years old to remove cases potentially delivered by residents or fellows, since junior physicians may introduce distinct team dynamics related to hierarchical supervision, rather than subtler gender-based dynamics (Column 3). Similarly, we show that results are robust to excluding deliveries occurring at teaching hospitals to account for hierarchical team structures between "true" attendings (defined by academic rank) and residents (Column 4). Finally, results are robust to excluding all hospitals without at least two male and female physicians in the data *and* at least one instance of each team type, to ensure that results are not driven by unusual weighting of estimates at hospitals where each of the four types of gender-mixed teams either can not, or do not, form (Column 5).

Varying outcome definitions: In Appendix Table A.4 we show that results are robust to both more expansive and restrictive definitions of maternal complications experienced during labor and delivery. First, to account for changes from ICD-9 to ICD-10 in October 2015 that could potentially influence how maternal complications are coded, we show robustness to only including quarter-years 2006-2015Q3 (Column 1). Second, while we use the standard CDC-defined metric for severe maternal morbidity as our main measure, OB-GYNs have proposed several alternative measures (Bateman et al., 2013; Snowden et al., 2021). As expected, we find larger effect sizes when using a more expansive measure (Column 2) and smaller effect sizes when using a more restrictive measure (Column 3).

Including team or physician fixed effects: Our main specification in Equation 1 allows us to make comparisons *across* all four team types while extensively controlling for patient and physician characteristics as well as time and hospital fixed effects. However, to better understand gender dynamics depending on which physician is in the Lead or Assisting role, we employ more precise tests using team or physician fixed effects (see Appendix Section D.2 for specification details). First, we limit the sample to mixed-gender teams and include a fixed effect for the physician team, irrespective of the physician's role in the team. For example, if Drs. Bob and Mary work together, including a fixed effect for the Bob-Mary team would isolate differences in outcomes when Bob is the Lead and Mary is the Assisting versus when Mary is the Lead and Bob is the Assisting. As seen in Appendix Figure A.2 Subfigure (a), we find that within the same mixed-gender team, maternal complications are higher when the Lead is a female than when the Lead is a male, consistent with the results from our main specification (Figure I) and now the difference is statistically significant at the 5% level. This analysis suggests that mixed-gender teams perform worse when a female physician is in the Lead role (a result we explore in greater detail as part of our mechanisms).

Second, we estimate a within-physician regression that includes a physician fixed effect for the Lead physician as the Lead is informally in charge of the entire encounter. For this analysis, we split the sample by the gender of the Lead physician (as the gender of the Lead cannot vary within a physician) and estimate how outcomes differ when the same Lead physician is paired with an Assisting of the same or opposite gender. In other words, it allows us to statistically test differences between F_L–F_A and F_L–M_A teams, and separately, between M_L–M_A and M_L–F_A teams. Additionally, including Lead physician fixed effects accounts for the Lead physician's time-invariant preferences (e.g., for certain schedules, types of obstetric procedures, or for specific Assisting physicians, should more than one be available) and abilities (e.g., to work effectively in a team) that could influence patient outcomes.[9]

[9] As discussed in Section 2.1 and Appendix Section B, the Lead physician is temporally the first physician assigned to the patient *and* also responsible for the entire clinical encounter. Therefore, assigning a fixed effect to the Lead physician naturally aligns with their central role and primary responsibility for patient outcomes.

In Appendix Figure A.2 Subfigure (b), we show that results closely match our main specification: maternal complication rates are significantly lower when the same female Lead physician is in a same-gender team (F_L–F_A) compared to mixed-gender team (F_L–M_A), whereas maternal complication rates are significantly higher when the same male Lead physician is in a same-gender (M_L–M_A) compared to mixed-gender team (M_L–F_A).

3.3 Supporting the identifying assumption: Can the relationship between gender mix and maternal complications be interpreted causally?

In Section 2.3, we provide conceptual support for our identification strategy, which assumes that patients are assigned quasi-randomly to the gender mix of the team. (We emphasize that patients do not need to be quasi-randomly assigned to *teams*, which is a much more stringent assumption). In Section 3.2, we show that our main result remains robust to multiple specifications, variable definitions, and samples, which lends credence to our identifying assumption. Now, we provide direct empirical evidence in support of this assumption, which simultaneously helps rule out some of the most obvious alternative explanations.

There are two primary alternative explanations for our findings. First, teams with different gender mixes may systematically differ in the types of patients they care for. For example, if male-only teams disproportionately treat higher-risk patients compared to female-only teams, they would naturally exhibit higher complication rates. Second, teams of different gender mixes may form endogenously along dimensions such as familiarity or skill, that independently influence performance. For example, if female Lead physicians preferentially select Assisting physicians with whom they are more familiar and who are themselves more likely to be female, this familiarity could improve maternal outcomes. While our main specification controls for patient risk factors, the age and experience of both the Lead and Assisting physician, and familiarity between physicians, we provide additional evidence that these two alternative channels do not explain why female-only teams outperform male-only teams, yet female-led mixed-gender teams perform worse than male-led ones.

3.3.1 Does patient risk of complication differ by team gender mix?

Consistent with prior literature, we identify 23 maternal risk factors observable *before the onset of labor*[10] that can complicate childbirth and increase the likelihood of maternal complications—such as diabetes, hypertension, and history of C-section (the full list of risk factors is provided in Footnote 8). Appendix Figure A.3 Subfigure (a) plots the raw (unadjusted) relationship between the number of these risk factors and the average rate of maternal complications, showing a clear positive association. For instance, over 25% of patients with 10 or more risk factors experience a maternal complication in our sample. Using only these observable risk factors, we estimate each patient's predicted probability of experiencing a maternal complication (denoted by $\widehat{MC_i^{RF}}$).[11] In Appendix Figure A.3 Subfigure (b) we show that this predicted measure closely tracks actual maternal complications observed in our data. We thus establish that these 23 risk factors, and $\widehat{MC_i^{RF}}$ (the predicted likelihood of maternal complication based on these risk factors) are appropriate measures of patient risk and can be used to assess whether patient risk differs across the four team types.

[10] We focus specifically on risk factors observable *prior* to labor onset because these factors can be known by the patient and physician before team formation, potentially influencing team selection.

[11] We generate these predicted probabilities ($\widehat{MC_i^{RF}}$) by regressing a maternal complication indicator on indicators for the 23 risk factors and the number of diagnosis codes recorded in the patient encounter. See Appendix Section E for more details.

Figure II Panel A shows that patient risk does not vary by a team's gender mix. Across the four team types, the distribution of the raw number of risk factors observed per patient is the same (Subfigure (a)) and the means of $\widehat{MC_i^{RF}}$ are comparable (Subfigure (b)). We also show that the distribution of $\widehat{MC_i^{RF}}$ are very similar across the four team types (Appendix Figure A.5), altogether providing compelling evidence that the teams are treating patient populations of similar risk.

3.3.2 Are physician teams being formed endogenously?

Team formation by gender: Another concern is that teams are formed endogenously according to gender (i.e., Lead physicians are more likely to select into teams with Assistings of a certain gender). To directly test whether this is the case, we regress a binary indicator for the Lead physician's gender on a binary indicator for the Assisting physician's gender (1 if female, 0 if male), with similar controls as in Equation 1. We plot the point estimates of this regression in Figure II Panel B Subfigure (c). We show that the gender of the Lead physician does not predict the gender of the Assisting physician. That is, female and male Leads are just as likely to work with a female (or, equivalently, male) Assisting physician. This result holds consistently for (i) high-risk *and* low-risk patient groups (where high-risk patients have more than the median number of risk factors), as well as for (ii) high-familiarity *and* low-familiarity teams (where high-familiarity teams are defined as those with cumulative joint deliveries above the sample median). Given these results, it is unlikely that a male (or female) Lead physician is systematically selecting an Assisting physician based on their gender, or does so specifically for high-risk patients, or based on prior familiarity. These results suggest that the endogenous formation of teams by gender is unlikely to be a concern in this setting.

Team formation under more random and less random conditions: We cannot test every potential dimension on which teams might form endogenously, since many dimensions may be unknown or unobservable. However, we can provide conceptual "bounds" for identification. Specifically, hospitals with fewer physicians on call *and* higher workloads – i.e., hospitals with "more random" conditions – likely have less opportunity for endogenous team formation, making quasi-random assignment more credible. Under these conditions, a Lead physician is effectively "forced" to work with whichever Assisting physician is next available, as there may be no alternative available at the time. Conversely, hospitals with more physicians and lower workloads – i.e., those with "less random" conditions – may have more organizational slack, creating opportunities for endogenous team formation. In Figure II Panel B Subfigure (d) we show that our main result holds when the sample is restricted to hospitals with more random conditions (those with *below*-median number of unique physicians per quarter and an *above*-median patient-to-provider ratio, ($\beta_3 = -.0076$, p-val < 0.001)) as well as when restricted to hospitals with less random conditions (those with an *above*-median number of unique physicians per quarter and a *below*-median patient-to-provider ratio, ($\beta_3 = -.0042$, p-val < 0.005)). While these two types of hospitals are clearly different and have different outcomes, these results provide further confirmation that the endogenous formation of teams is unlikely to be driving our results.

Team formation by skill: Lastly, we consider whether teams may be forming according to physician skill in ways correlated with gender. To do so, we first construct a proxy for individual physician skill using the solo sample of births (deliveries with only one physician on the discharge record, since these outcomes can be more directly attributed to an individual physician's skill). Specifically, we regress maternal complications on the 23 patient risk factors, hospital-year and quarter-year fixed effects, and physician fixed effects (details in Appendix E). The physician fixed effects serve as a skill

proxy because they capture the physician-specific component of maternal complications that cannot be explained by patient clinical risk or institutional practices varying by hospital-year or quarter-year. Appendix Figure A.6 Subfigure (a) shows the distribution of these skill proxies and confirms similar skill distributions among male and female physicians. Then, as a conceptual validation of this skill proxy, in Appendix Figure A.6 Subfigure (b), we show that teams with lower-skilled physicians have worse maternal complication rates compared to teams with higher-skilled physicians.

Given the inherent noisiness of physician fixed-effect estimates, we rank physicians based on their fixed effects (normalized from 0 to 1), where a rank of 0 corresponds to the highest skill and a rank of 1 to the lowest skill. In Appendix Figure A.6 Subfigure (c), we take the average of the rank of the Lead and Assisting physician, and find that this average is close to 0.5 for all four team types, indicating similar average skill levels across all teams. This figure also shows that the difference in skill ranks between Lead and Assisting physicians within each team type is consistently small and comparable across team types. This finding rules out several alternative explanations – for example, that male-only teams have systematically lower average skill, or that female Leads in F_L–M_A teams consistently call in higher-skilled male Assistings. Overall, we find limited evidence that teams form based on physician skill in ways correlated with team gender mix.

3.4 Heterogeneities in team performance

Heterogeneity by patient risk: A patient's medical complexity can influence both maternal complication rates and how teams are formed. For example, a team may form more spontaneously for a complication arising during a low-risk birth, while the labor and delivery ward may be alerted to and plan for a high-risk birth.[12] To test if one of these scenarios is driving our results, we split the sample into patients with above median count of clinical risk factors (defined as high-risk) and below median count of clinical risk factors (defined as low-risk) and re-estimate Equation 1 for each subsample. As expected, in Figure III Subfigure (a), we find that maternal complication rates are almost four times higher for high-risk compared to low-risk patients. However, the main takeaways are unchanged: female physicians (relative to male physicians) have lower maternal complication rates in same-gender teams for both low-risk and high-risk patients. Female-led mixed-gender teams (F_L–M_A) also perform worse than male-led mixed-gender teams (M_L–F_A) when treating low-risk patients.

Heterogeneity by patient race: A large body of research has documented persistent racial and socioeconomic disparities in maternal outcomes, especially for Black women (Alsan & Yearby, 2024; Hill, Artiga, & Published, 2022; Hill et al., 2024; Kennedy-Moulton et al., 2022). Such disparities are widely regarded as indicators of lower-quality obstetric care (Howell & Zeitlin, 2017). Therefore, we examine whether racial disparities in maternal complications exist by team type. In Figure III Subfigure (b), we find that female-only teams are unique in two ways: they have the lowest maternal complication rates for Black women, and they have no differences in complication rates between Black and non-Black women. No other team type achieves either of these outcomes. Male-only teams, in fact, exhibit the largest racial disparity and have the worst outcomes for Black patients out of all four team types. Overall, regardless of patient race, we continue to observe the main pattern that female-only teams (F_L–F_A) have lower maternal complication rates than male-only teams (M_L–M_A), but in mixed-gender teams, female-led teams (F_L–M_A) have higher complication rates than male-led teams (M_L–F_A).

[12] High-risk births are not automatically assigned to teams. For example, a patient may be over 35 and have a breech baby with an indication for C-section, but only have one OB-GYN involved in labor and delivery.

Heterogeneity by patient insurance status: Whether a patient is publicly or privately insured could potentially influence the type of treatment they receive and the type of physician who treats them. For example, Medicaid typically provides lower reimbursement for labor and delivery than privately insured patients, especially for C-sections. Extensive research has shown that OB-GYNs respond to these different financial incentives (Foo, Lee, & Fong, 2017; Gruber & Owings, 1996; Johnson & Rehavi, 2016), and so to the extent that male or female OB-GYNs respond differently to financial incentives, this could influence team formation or dynamics in ways that influence maternal complications. However, as seen in Figure III Subfigure (c), we do not find statistically different maternal complication rates between privately insured and publicly insured patients within each team type, and the key performance patterns across the four team types are similar to our main results.

4 MECHANISMS

In the previous section, we provide evidence that team gender mix causally affects a patient's likelihood of maternal complications. In this section, we shed light on potential mechanisms explaining these results by investigating two questions: (1) *How* does team gender mix affect team performance, and (2) *Why* does team gender mix affect team performance?

4.1 How does gender mix affect team performance?

How might the team's gender mix causally affect performance (i.e., maternal complication rates)? There are two distinct pathways through which this could occur. The first pathway, which we call the "Individual Preference" effect, assumes that men and women have inherently different preferences regarding clinical decisions, and thus, placing them in same- or mixed-gender teams will almost tautologically produce different outcomes, since the average preferred decision varies across teams with different gender mixes. For example, if female physicians prefer Choice A on average, and male physicians prefer Choice B on average, then female-only teams will choose Choice A and male-only teams will choose Choice B on average, and mixed-gender teams will choose some average of Choice A and Choice B. The second pathway, which we call the "Gender Mix" effect, refers to the direct effect of the team's gender mix itself. This pathway implies that, keeping the two members of the team the same but only changing their genders, could alter the team's decisions and outcomes.

Distinguishing clearly between these pathways is important, as each carries distinct policy implications. If gender mix affects team outcomes through the "Individual Preference" pathway, then a policymaker could change team decisions and outcomes by changing member preferences, either through incentives, information, or targeted behavioral interventions. If gender mix affects performance through the "Gender Mix" pathway, then improving team performance is more difficult, as gender is not mutable as a policy recommendation.

4.1.1 Operationalizing key constructs

Figure IV provides an intuitive graphic for these two pathways and outlines how we test them. At a high level, each physician has preferences over clinical decisions, and in a team, they incorporate these individual preferences to reach a team decision (Box A → Box B). These decisions then affect the team's performance (Box B → Box C). The gender of the team members can either affect team performance *indirectly* (through individual member preferences, Pathway 1), or *directly* (Pathway 2). See Appendix F.1 for more details on variable construction and analyses discussed in this section.

The key clinical decision that we focus on is whether to deliver the baby via vaginal birth or C-section. We focus on the C-section decision because it is widely recognized as a *discretionary* procedure with known risks, and is almost certainly overused (e.g., the average C-section rate is 41% in our data, while the World Health Organization recommends a target of 19%). Discretionary practices refer to medical interventions that are often left to physician judgment, frequently deviate from established best practices, and may introduce avoidable risks and complications into patient care (Gardella et al., 2001; Hartmann et al., 2005; Souza et al., 2010). Such discretion in C-section use contributes to the significant variation in C-section rates from physician to physician – independent of patient medical need (Allin et al., 2015; Epstein & Nicholson, 2009) – despite robust quasi-experimental evidence showing that unnecessary C-sections can be harmful to both mother and baby (Costa-Ramón et al., 2018; Halla et al., 2016; Tonei, 2019; Yu et al., 2023). It is precisely this discretionary nature that makes the team's C-section decision especially relevant and interesting in our setting, as it often requires joint decision-making – potentially reconciling differing preferences within the team – about whether or not to perform the procedure. As a robustness check, we also examine the use of C-sections on the subset of births that are defined as low-risk (live babies born at or beyond 37 weeks of gestation to women with no prior C-section, are singleton, and in the vertex presentation), as C-sections performed on low-risk patients may capture a smaller, but more certain, subset of inappropriate decisions than the use of C-sections broadly writ. We emphasize that C-sections are just *one* observable decision among the *many* unobservable clinical and non-clinical decisions that teams make. Thus, while we focus on the C-section decision to understand gender-driven team dynamics, the mechanisms we uncover likely generalize beyond C-sections to other discretionary team decisions.

We now describe how we measure each of our constructs in Figure IV.

1. Individual member preferences (Box A): The discretionary nature of C-sections often leads to variation in physicians' preferences for performing them. Thus, a physician's C-section preference can be viewed as their practice style or internal threshold for choosing a C-section, independent of external circumstances. This perspective was echoed in our conversations with OB-GYNs; for instance, one OB-GYN explained: "You come to know who has what preferences and tolerance of when to [perform] a C-section [during a delivery]." We proxy for each physician's individual preference for C-sections using the solo sample, where we isolate each physician's fixed effect (FE) from regressing an indicator for whether a patient receives a C-section, on physician FEs, patient controls, and quarter-year and hospital-year FEs as in Equation 1. This physician FE captures how much a patient's likelihood of having a C-section changes if their delivery is performed by *Physician A* instead of *Physician B*, while keeping constant patient, hospital, and time characteristics. We repeat the same estimation to obtain the fixed effect for a two-member physician team – i.e., the team preference for C-section use – but using the team sample instead of the solo sample.

   First, we show in Appendix Figure A.7 Panel A that stronger preferences for C-section – whether they be of the individual physician or the team – are associated with higher maternal complication rates for all births and low-risk births. Second, in Appendix A.7 Figure Panel B we show that male physicians have stronger preferences for C-sections than female physicians, which makes it important to distinguish between the effect of preferences and the effect of team gender mix on outcomes.[13]

2. Team decision (Box B): We examine the team decision to perform a C-section during the delivery.

[13] Even though male physicians perform more C-sections, we do not find differences in physician skill by gender in the solo sample as explained in Section 3.3.2. One possible reason may be because of patient selection: individual male and female physicians may be able to select the panel of patients they see in ways that teams cannot.

We show in Appendix Figure A.7 Panel C that patterns of team decisions to perform a C-section closely align with the key patterns of maternal complications in Figure I, supporting our assumption of a non-zero causal effect of C-sections on maternal complications.[14] As further support, we find the same pattern exists on the subset of low-risk births.

3. Team performance (Box C): We measure team performance as we have in the paper thus far, using maternal complications.

Following the framework in Figure IV, we now test whether individual member preferences for C-section (Box A) translate into meaningful differences in team decisions (Box B) and, therefore, team performance (Box C) by estimating the following linear probability model:

$$\begin{aligned} y_{ijkht} = \boldsymbol{\beta}(Lead_pref_csec_j \times Asst_pref_csec_k) + \eta Diff_in_prefs \\ + \gamma X_i + \delta Z_j + \phi T_k + \alpha_t + \alpha_{hy} + \varepsilon_{ijkht} \end{aligned} \tag{2}$$

where y is the outcome of interest – indicator for C-section or indicator for maternal complications – for patient i under the care of Lead physician j and Assisting physician k in hospital h and quarter-year t. $Lead_pref_csec_j$ and $Asst_pref_csec_k$ capture the C-section preferences of the Lead and Assisting, respectively. Given that fixed effects are noisily estimated, we convert the continuous fixed effect into a binary measure: a physician is classified as having a preference for C-sections if their C-section fixed effect estimate is above 0 (the sample's average C-section preference) and a preference for vaginal deliveries if it is below 0.[15] $Diff_in_prefs$ controls for the continuous difference between the Lead and Assisting physicians' fixed effects to account for how close or far apart the two physicians' preferences are. All other control variables and fixed effects are as in Equation 1. As robustness checks, we offer two additional ways of defining preferences, where we divide the continuous FEs into (i) terciles: "preferences for C-section", "neutral preferences", and "preferences for vaginal deliveries"; or (ii) quintiles: "strong preferences for C-section" "weak preferences for C-section", "neutral preferences", "weak preferences for vaginal deliveries", and "strong preferences for vaginal deliveries."

Figure V Subfigures (a) and (b) show the fitted values obtained from Equation 2 for C-sections and maternal complications, respectively. On average, we find that teams are most likely to perform a C-section – and incur a maternal complication – when both physicians prefer C-sections, and least likely when both physicians prefer vaginal deliveries. Teams consisting of one physician who prefers C-sections and one who prefers vaginal deliveries have C-section decision rates *and* maternal complication rates that fall between these two extremes. This intermediate pattern observed when physicians have conflicting C-section preferences suggests some type of joint decision-making within teams, rather than simply relying on the "common" preference, as is the case when the two physicians' preferences align. Overall, our Figure IV framework appears valid: individual member preferences for C-section influence team decisions, which in turn affect maternal complications (Box A → Box B → Box C).

4.1.2 The "Individual Preference" effect vs the "Gender Mix" effect

If the Individual Preference effect (Pathway 1) fully explains variation in team performance, then holding each individual physician's preferences constant, the gender mix of the team should not affect

[14] We remain agnostic about the precise causal effect of C-section use on maternal complications, as estimating this magnitude is not central to our analysis. Instead, we focus on how C-section use varies *across team types*, because if C-sections cause maternal complications at all, team gender mix could influence complication rates via differential C-section use. Crucially, any endogeneity between C-section use and complications should remain constant across team types, given the similarity in patient risk profiles.

[15] Using this categorization, 817 physicians are identified as having a preference for C-sections, and they perform on average, 1016 deliveries each in our data, with a 47.5% C-section rate. Conversely, 1,227 physicians are identified as having a preference for vaginal deliveries, and they perform an average of 901 deliveries each, with a 35.5% C-section rate.

team decisions or performance. That is, the Individual Preference effect implies that *within* any given combination of preferences, all four team types should have the same C-section rate. However, if the Gender Mix effect (Pathway 2) is operative, differences in C-section rates should persist across the four team types even *within* the same combination of individual preferences. To test this, we decompose each estimate shown in Figure V Subfigures (a) and (b) into four separate estimates for each team type. Specifically, we estimate a modified version of Equation 2, in which we include a triple interaction between $Lead_pref_csec_j \times Asst_pref_csec_k \times Team_{jk}$, where $Team_{jk}$ identifies each of the four team types.

We provide clear evidence in Figure V Subfigures (c) and (d) that the Gender Mix effect exists: even within a combination of physician preferences, there are meaningful differences in both team decisions and performance by team gender mix. At the extreme, our results suggest that if one kept everything about the two physicians in the team exactly the same – but simply changed one or more physician's gender – the team's clinical decisions and subsequent patient outcomes would change substantially. This striking result highlights the distinct, direct role that physician gender plays in this setting. We highlight two key patterns below, which mirror those in our main Figure I.

First, among same-gender teams, male-only teams are consistently more likely to perform C-sections than female-only teams, regardless of whether the preferences of the two team members conflict or align (Figure V Subfigure (c)). This translates into higher maternal complication rates for male-only teams, while female-only teams generally have the lowest complication rates (Figure V Subfigure (d)). Interpreted differently: even if both physicians have identical preferences for C-section, a patient is more likely to have a C-section and more likely to experience a maternal complication under the care of the male-only team than under a female-only team.

Second, among mixed-gender teams, female-led teams (F_L–M_A) are more likely to perform C-sections than male-led teams (M_L–F_A) (Figure V Subfigure (c)), but this tendency translates into worse maternal outcomes *only* when there is a conflict in preferences between team members (Figure V Subfigure (d)). We highlight this result because the pattern – where F_L–M_A teams make different decisions and experience worse outcomes than M_L–F_A teams – is one of our most consistent findings and serves as a key indicator for understanding *why* gender mix affects team performance. This result suggests that team decision-making may be more straightforward when preferences align, but becomes more complex when preferences differ, especially when a female physician is in the Lead role. However, conflict in preferences should not always harm performance nor should alignment always be beneficial to outcomes, as incorporating differing perspectives can lead to higher quality decision-making. In the next section, we show that conflict can enhance performance if the final decision effectively integrates and leverages the strengths of both perspectives.

As robustness checks, we confirm that our key result – that team gender mix impacts outcomes even after holding each physician's preferences constant – is robust to alternative definitions of physician preferences. Specifically, we estimate a modified version of Equation 2 that replaces the binary preference variable with categorical variables defined by terciles or quintiles of preferences. Appendix Figure A.8 compares outcomes between physicians at the extremes of these categories (strongest preferences for C-sections vs. vaginal deliveries), showing similar results to Figure V. Notably, the further apart the individual physicians' preferences, the greater the influence of team gender mix on decisions and outcomes becomes.

4.2 Why does gender mix affect team performance?

Figure V demonstrates that the team's gender mix plays a direct and causal role in shaping both team decision-making and performance. In this section, we investigate why this may be the case. We hypothesize that gender mix introduces social norms around gender (i.e., gender norms), which can affect how effectively the team communicates and collaborates.

While we cannot test for the presence of gender norms directly, we can make inferences about them from patterns in the data. We find two patterns that support this hypothesis. First, we find that gender mix affects how team-members' preferences are incorporated into team decisions. That is, same-gender teams are more likely to make decisions that reflect team members' average preference than are mixed-gender teams, and female-only teams, in particular, do not default towards more discretionary procedures such as C-sections. Second, we find that gender mix affects how resilient teams are to challenging conditions, with female-led mixed-gender teams performing especially poorly under these conditions, potentially because female leaders invert traditional gender norms around leadership. We discuss each of these findings separately below.

4.2.1 Gender mix affects the incorporation of individual preferences into team decisions.

Gender norms may determine whose preferences in a team carry more weight in the final decision; it could push individual preferences in a particular direction due to peer pressure or social desirability bias; it may affect how preferences are expressed or elicited within the team; and it could alter collective risk preferences. For example, the team may opt for a more (or less) risky choice than its individual members would choose on their own.

To understand how individual preferences are incorporated into the team's C-section decision, we plot three estimates in Figure VI for each team type. First, we plot each team's *joint* C-section preferences (blue bars).[16] Positive values imply that the team jointly prefers C-sections while negative values imply that the team jointly prefers vaginal deliveries. Second, we plot the average of the *individual* physician preferences of the two team members for each team (red bars). Positive values imply that the team members, on average, prefer C-sections when delivering alone, while negative values imply that the team members, on average, prefer vaginal deliveries when delivering alone. Third, we plot the difference between these two estimates (gray bars). Positive values for the gray bars imply that the team has a stronger preference for C-sections than the team members do when they deliver alone; negative values imply that the team has a stronger preference for vaginal deliveries than the team members do when they deliver alone.

Figure VI shows that same-gender teams' joint preferences align closely with their members' individual preferences (the red and gray bars move in the same direction), while mixed-gender teams display notable divergence between individual and joint preferences (the red and gray bars move in opposite directions). In other words, same-gender teams more closely reflect the average preferences of their members in their joint decisions than do mixed-gender teams. This result suggests that same-gender teams are more skilled at aggregating individual preferences – whether aligned or conflicting – into cohesive team decisions, potentially because of greater ease of collaboration.

This figure also shows that female-only teams are the *only* team types that do not default towards more C-sections than their members' individual preferences would suggest (i.e., they are the

[16] A team's joint C-section preference measures how much a patient's likelihood of receiving a C-section changes depending on whether Team A or Team B performs the delivery. We isolate team fixed effects by regressing a C-section indicator on patient, hospital, and time controls. To ensure reliable estimates of team fixed effects, we limit our analysis to teams that have performed at least 35 deliveries in our data.

only team type with a negative grey bar). In fact, they perform more vaginal deliveries than their individual preferences would suggest, which could explain female-only teams' superior performance across all four team types. In contrast, male-only teams not only default toward performing more C-sections as a team than they would individually, but they are also the only team type whose members, on average, have a baseline preference for C-sections (i.e., the only team type with a positive red bar)..

Why would all teams, except female-only teams, default towards more discretionary practices like C-sections? One reason is that the decision to perform a C-section has often been described as the "path of least resistance" (Main et al., 2011), especially when there is conflict in preferences. For example, White VanGompel et al. (2019) found that discordant attitudes between nurses and physicians over birth practices were associated with higher C-section rates. In our setting, teams may default to performing C-sections to overcome disagreements about the optimal course of action – or even simply because one member of the team strongly favors a C-section.

Another reason is that team settings may exacerbate gender-based behavioral differences. Men exhibit higher risk tolerance and greater action orientation than women, traits that are further amplified in team settings (Lamiraud & Vranceanu, 2018). Additionally, women's opinions and expertise are often undervalued, making them less likely to voice their views or see those views reflected in final decisions, especially in male-dominated settings (Coffman, Flikkema, & Shurchkov, 2021; Thomas-Hunt & Phillips, 2004). Taken together, merely including one male physician on a team – regardless of whether he prefers vaginal or C-section deliveries – may push the team towards performing more C-sections, as this choice is both more action-oriented and, riskier when medically unnecessary. This is apparent in Figure V Panel B, which shows that in almost all instances, teams with at least one male physician perform more C-sections than teams with no male physicians, regardless of whether that male physician(s) prefers C-sections or not. Overall, these results are consistent with gender norms shaping how individual preferences are aggregated into team decisions in ways that particularly harm the performance of male-only teams but benefit female-only teams.

4.2.2 Gender mix affects team resilience to challenges.

Gender norms may also affect how team members communicate, solve conflicts, and view leaders. These dynamics can shape how effectively teams navigate challenges and make decisions, ultimately influencing performance. While we cannot directly observe these complex social dynamics in our data, we can examine team performance under challenging conditions and consider the role of gender norms in shaping how resilient teams are to these challenges. For example, a consistent pattern in our findings is that female-led mixed-gender teams (F_L–M_A) make different decisions and perform worse than male-led mixed-gender teams (M_L–F_A), despite both having one male and one female physician. The key distinction between these team types is that a female Lead with a male Assisting runs counter to traditional gender norms around leadership (recall that in our sample, male physicians are the team lead in 58% of births). The inversion of these norms may introduce friction, making it harder for team members to perform when confronting additional challenges, ultimately leading to worse outcomes.

We thus examine team performance under four conditions that could make teamwork more challenging (variable construction described in Appendix F.2.2):

1. Conflict in preferences: When the Lead and Assisting physicians have conflicting preferences about C-section use, coordination and dispute resolution may become more difficult.
2. Lead being younger than Assisting: When the Lead physician is notably younger than the Assist-

ing physician, this age difference can disrupt hierarchy or create experiential tension, potentially undermining the Lead's authority (Buengeler, Homan, & Voelpel, 2016).

3. Limited prior collaboration: When physicians have rarely worked together before, they may be less familiar with each other's communication style and decision-making approach, hindering trust-building and smooth coordination (Chen, 2021).
4. High hospital strain: Performing a delivery when the hospital is near capacity can intensify teamwork challenges due to heightened time pressures, cognitive load, and resource scarcity (Singh & Venkataramani, 2022).

Figure VII (a) - (d) show the maternal complication rates for each team type, both with and without the presence of the four aforementioned challenges to teamwork. Across all four figures, we find that female-led mixed-gender teams (F_L–M_A) are the least resilient: they have consistently worse performance under these conditions than they do without. Specifically, F_L–M_A teams are the only team that exhibit higher maternal complication rates when the two team-members have conflicting C-section preferences (Subfigure (a)), when the Lead is younger than the Assisting (Subfigure (b)), when the Lead and Assisting have not worked together extensively before (Subfigure (c)), and when the Lead and Assisting are delivering a baby during a period of high hospital strain (Subfigure (d)). Interestingly, Figure VII Subfigure (a) shows that the other team types *improve* performance (i.e., have lower maternal complication rates) when their members have conflicting preferences. This result suggests that diverse perspectives may lead to better outcomes if they encourage more deliberative decision-making. See Appendix Section F.2.2 for more discussion.

Gender norms influencing team dynamics are the most likely explanation for this persistent pattern: F_L–M_A teams may struggle most because female leaders disrupt traditional gendered expectations around leadership. Prior research has extensively documented that female leaders face unique challenges. For example, female leaders are often viewed less favorably than men (Eagly & Karau, 2002), perceived as less effective even with no performance differences (especially by male subordinates) (Beaman et al., 2009; Eagly, Makhijani, & Klonsky, 1992; Gangadharan et al., 2016), socially penalized for violating traditional feminine norms (Eagly & Johannesen-Schmidt, 2001; Rudman & Glick, 1999), and receive less credit or reward for team successes (Grossman et al., 2016). These findings suggest female-led mixed-gender teams perform worse when confronted with teamwork challenges because they face a unique social context that is inherently detrimental to teamwork.

5 CONCLUSION

This paper studies the causal effect of team gender mix on team performance in the high-stakes setting of childbirth. Childbirth is an important context because it involves critical, real-time decision-making, requires extensive teamwork under pressure, and features clear and consequential measures of performance in the form of maternal complications. We find that female-only physician teams have the lowest rates of maternal complications and male-only teams have the highest, but within mixed-gender teams, female-led teams perform worse than male-led teams. These performance differences cannot be explained by differences in patient populations or by teams forming based on gender, familiarity, or skill. Instead, we provide evidence that differences in maternal complication rates across team types can be directly attributed to the gender mix of the team and that gender norms appear to significantly shape team decision-making. In particular, gender mix can affect how individual preferences are incorporated into team decisions as well as how resilient teams are to working under

potentially challenging conditions.

We highlight two main limitations of our study. First, because all patients are female, we cannot assess whether gender concordance between patients and physician teams contributes to outcomes. Female-only teams may outperform others due to gender "match" effects, such as greater patient comfort in expressing their preferences or provider attentiveness to women's health. However, concordance alone is unlikely to explain our findings. If it were the main driver, then having a single woman on the team should consistently have better outcomes than male-only teams, which is not the case for female-led mixed-gender teams. Moreover, if gender concordance improved maternal outcomes, then female physicians delivering individually should consistently outperform male physicians – but we do not find this to be the case in the solo sample. Second, our analyses rely on administrative data, which does not provide the necessary detail to more precisely explore mechanisms that may be responsible for differential patient outcomes by team gender mix. Future research using richer clinical data, such as electronic health records or physician notes, could provide more insight into how gender mix affects teamwork and decision-making.

Despite these limitations, our findings offer new insights for managing an increasingly diverse workforce. Organizations often attempt to shift individual decision-making through performance feedback, incentives, or guidelines, but our results suggest that the gender mix of the team can shape outcomes in ways that cannot be addressed by targeting individual behavior alone. Instead, organizations could consider training programs and team-building exercises that promote inclusive leadership, address implicit gender norms around authority, and enhance trust and communication in teams to help reduce gender-based frictions (Carnes et al., 2015; Castro, Englmaier, & Guadalupe, 2024). Such approaches may be especially important in male-dominated expert professions like medicine, where traditional hierarchies and gender expectations may hinder effective collaboration. Further research is needed to evaluate the extent to which these findings apply in other professional settings and to identify concrete strategies for enhancing teamwork in increasingly gender-diverse environments.

References

ACOG. (2012). Communication Strategies for Patient Handoffs. Retrieved June 11, 2024, from https://www.acog.org/clinical/clinical-guidance/committee-opinion/articles/2012/02/communication-strategies-for-patient-handoffs

ACOG. (2016a, February). The Obstetric and Gynecologic Hospitalist, Committee Opinion Number 657. Retrieved May 17, 2025, from https://www.acog.org/clinical/clinical-guidance/committee-opinion/articles/2016/02/the-obstetric-and-gynecologic-hospitalist

ACOG. (2016b). Severe maternal morbidity: Screening and review. American Journal of Obstetrics and Gynecology, 215(3), B17–22. https://doi.org/10.1016/j.ajog.2016.07.050

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. Economics Letters, 80(1), 123–129. https://doi.org/10.1016/S0165-1765(03)00032-6

Allin, S., Baker, M., Isabelle, M., & Stabile, M. (2015, March). Accounting for the Rise in C-sections: Evidence from Population Level Data. https://doi.org/10.3386/w21022

Alsan, M., & Yearby, R. (2024). Health Equity in the 2024 U.S. Presidential Election [Publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMp2410598]. New England Journal of Medicine, 391(15), 1374–1377. https://doi.org/10.1056/NEJMp2410598

Aparicio Fenoll, A., & Zaccagni, S. (2022). Gender mix and team performance: Differences between exogenously and endogenously formed teams. Labour Economics, 79, 102269. https://doi.org/10.1016/j.labeco.2022.102269

Apesteguia, J., Azmat, G., & Iriberri, N. (2012). The Impact of Gender Composition on Team Performance and Decision Making: Evidence from the Field [Publisher: INFORMS]. Management Science, 58(1), 78–93. https://doi.org/10.1287/mnsc.1110.1348

Bateman, B. T., Mhyre, J. M., Hernandez-Diaz, S., Huybrechts, K. F., Fischer, M. A., Creanga, A. A., Callaghan, W. M., & Gagne, J. J. (2013). Development of a comorbidity index for use in obstetric patients. Obstetrics and Gynecology, 122(5), 957–965. https://doi.org/10.1097/AOG.0b013e3182a603bb

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., & Topalova, P. (2009). Powerful Women: Does Exposure Reduce Bias?*. The Quarterly Journal of Economics, 124(4), 1497–1540. https://doi.org/10.1162/qjec.2009.124.4.1497

Bertrand, M., Black, S. E., Jensen, S., & Lleras-Muney, A. (2014, June). Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labor Market Outcomes in Norway. https://doi.org/10.3386/w20256

Born, A., Ranehill, E., & Sandberg, A. (2022). Gender and Willingness to Lead: Does the Gender Composition of Teams Matter? The Review of Economics and Statistics, 104(2), 259–275. https://doi.org/10.1162/rest_a_00955

Boyle, P. (2021, February). Nation's physician workforce evolves: More women, a bit older, and toward different specialties. Retrieved August 22, 2024, from https://www.aamc.org/news/nation-s-physician-workforce-evolves-more-women-bit-older-and-toward-different-specialties

Buengeler, C., Homan, A. C., & Voelpel, S. C. (2016). The challenge of being a young manager: The effects of contingent reward and participative leadership on team-level turnover depend on leader age [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/job.2101]. Journal of Organizational Behavior, 37(8), 1224–1245. https://doi.org/10.1002/job.2101

Carnes, M., Devine, P. G., Manwell, L. B., Byars-Winston, A., Fine, E., Ford, C. E., Forscher, P., Isaac, C., Kaatz, A., Magua, W., Palta, M., & Sheridan, J. (2015). Effect of an Intervention to Break the Gender Bias Habit for Faculty at One Institution: A Cluster Randomized, Controlled Trial. Academic medicine : journal of the Association of American Medical Colleges, 90(2), 221–230. https://doi.org/10.1097/ACM.0000000000000552

Castro, S., Englmaier, F., & Guadalupe, M. (2024). Fostering Psychological Safety in Teams: Evidence from an RCT [Publisher: Academy of Management]. Academy of Management Proceedings, 2024(1), 16624. https://doi.org/10.5465/AMPROC.2024.16624abstract

CDC. (2024, May). Identifying Severe Maternal Morbidity (SMM). Retrieved June 11, 2024, from https://www.cdc.gov/maternal-infant-health/php/severe-maternal-morbidity/icd.html

Chan, D. C. (2021). Influence and information in team decisions: Evidence from medical residency [Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2425]. American Economic Journal: Economic Policy, 13(1), 106–137. Retrieved August 1, 2025, from https://www.aeaweb.org/articles?id=10.1257/pol.20180501

Chen, Y. (2021). Team-Specific Human Capital and Team Performance: Evidence from Doctors. American Economic Review, 111(12), 3923–3962. https://doi.org/10.1257/aer.20201238

Coffman, K., Flikkema, C. B., & Shurchkov, O. (2021). Gender stereotypes in deliberation and team decisions [Publisher: Elsevier]. Games and Economic Behavior, 129(100), 329–349. Retrieved March 23, 2025, from https://ideas.repec.org//a/eee/gamebe/v129y2021icp329-349.html

Corredor-Waldron, A., Currie, J., & Schnell, M. (2024, August). Drivers of Racial Differences in C-Sections. https://doi.org/10.3386/w32891

Costa-Ramón, A. M., Rodríguez-González, A., Serra-Burriel, M., & Campillo-Artero, C. (2018). It's about time: Cesarean sections and neonatal health. Journal of Health Economics, 59, 46–59. https://doi.org/10.1016/j.jhealeco.2018.03.004

Currie, J., & MacLeod, W. B. (2017). Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians. Journal of labor economics, 35(1), 18977. https://doi.org/10.3386/w18977

Eagly, A. H., & Johannesen-Schmidt, M. C. (2001). The leadership styles of women and men [Place: United Kingdom Publisher: Blackwell Publishing]. Journal of Social Issues, 57(4), 781–797. https://doi.org/10.1111/0022-4537.00241

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders [Place: US Publisher: American Psychological Association]. Psychological Review, 109(3), 573–598. https://doi.org/10.1037/0033-295X.109.3.573

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis [Place: US Publisher: American Psychological Association]. Psychological Bulletin, 111(1), 3–22. https://doi.org/10.1037/0033-2909.111.1.3

Epstein, A. J., & Nicholson, S. (2009). The formation and evolution of physician treatment styles: An application to cesarean sections. Journal of Health Economics, 28(6), 1126–1140. https://doi.org/10.1016/j.jhealeco.2009.08.003

Florida Legislature. (2023). 2023 Florida Statutes: Title XLV, Chapter 766, Section 766.301 - Legislative findings and intent. Retrieved June 6, 2025, from https://www.flsenate.gov/Laws/Statutes/2023/766.301

Foo, P. K., Lee, R. S., & Fong, K. (2017). Physician Prices, Hospital Prices, and Treatment Choice in Labor and Delivery [Publisher: The University of Chicago Press]. American Journal of Health Economics, 3(3), 422–453. https://doi.org/10.1162/ajhe_a_00083

Gangadharan, L., Jain, T., Maitra, P., & Vecci, J. (2016). Social identity and governance: The behavioral response to female leaders [Publisher: Elsevier]. European Economic Review, 90, 302–325. https://doi.org/10.1016/j.euroecorev.2016.01.003

Ganguli, I., Sheridan, B., Gray, J., Chernew, M., Rosenthal, M. B., & Neprash, H. (2020). Physician Work Hours and the Gender Pay Gap - Evidence from Primary Care. The New England Journal of Medicine, 383(14), 1349–1357. https://doi.org/10.1056/NEJMsa2013804

Gardella, C., Taylor, M., Benedetti, T., Hitti, J., & Critchlow, C. (2001). The effect of sequential use of vacuum and forceps for assisted vaginal delivery on neonatal and maternal outcomes. American Journal of Obstetrics and Gynecology, 185(4), 896–902. https://doi.org/10.1067/mob.2001.117309

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences*. The Quarterly Journal of Economics, 118(3), 1049–1074. https://doi.org/10.1162/00335530360698496

Goradia, D., & Chandrasekaran, A. (2024). Examining the role of single versus dual decision-making approach for patient care: Evidence from cardiology patients. J. Oper. Manag., 71(1), 11–39. https://doi.org/10.1002/joom.1340

Greenwood, B. N., Carnahan, S., & Huang, L. (2018). Patient-physician gender concordance and increased mortality among female heart attack patients. Proceedings of the National Academy of Sciences of the United States of America, 115(34), 8569–8574. https://doi.org/10.1073/pnas.1800097115

Gregory, K. D., Korst, L. M., Gornbein, J. A., & Platt, L. D. (2002). Using Administrative Data to Identify Indications for Elective Primary Cesarean Delivery. Health Services Research, 37(5), 1387–1401. https://doi.org/10.1111/1475-6773.10762

Grossman, P. J., Eckel, C., Komai, M., & Zhan, W. (2016). It Pays to Be a Man: Rewards for Leaders in a Coordination Game [Number: 38-16 Publisher: Monash University, Department of Economics]. Monash Economics Working Papers. Retrieved March 23, 2025, from https://ideas.repec.org//p/mos/moswps/2016-38.html

Gruber, J., & Owings, M. (1996). Physician financial incentives and cesarean section delivery. The Rand Journal of Economics, 27(1), 99–123.

Halla, M., Mayr, H., Pruckner, G. J., & Garcia-Gomez, P. (2016). Cutting Fertility? The Effect of Cesarean Deliveries on Subsequent Fertility and Maternal Labor Supply [Number: 9905

Publisher: Institute of Labor Economics (IZA)]. IZA Discussion Papers. Retrieved March 25, 2025, from https://ideas.repec.org//p/iza/izadps/dp9905.html

Hardt, D., Mayer, L., & Rincke, J. (2024). Who Does the Talking Here? The Impact of Gender Composition on Team Interactions [Publisher: INFORMS]. Management Science. https://doi.org/10.1287/mnsc.2023.03411

Hartmann, K., Viswanathan, M., Palmieri, R., Gartlehner, G., Thorp, J., & Lohr, K. N. (2005). Outcomes of Routine EpisiotomyA Systematic Review. JAMA, 293(17), 2141–2148. https://doi.org/10.1001/jama.293.17.2141

Henry, O. A., Gregory, K. D., Hobel, C. J., & Platt, L. D. (1995). Using ICD-9 codes to identify indications for primary and repeat cesarean sections: Agreement with clinical records. American Journal of Public Health, 85(8 Pt 1), 1143–1146. Retrieved May 24, 2024, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1615835/

Hill, L., Artiga, S., & Published, U. R. (2022, November). Racial Disparities in Maternal and Infant Health: Current Status and Efforts to Address Them. Retrieved September 14, 2024, from https://www.kff.org/racial-equity-and-health-policy/issue-brief/racial-disparities-in-maternal-and-infant-health-current-status-and-efforts-to-address-them/

Hill, L., Rao, A., Artiga, S., & Published, U. R. (2024, October). Racial Disparities in Maternal and Infant Health: Current Status and Efforts to Address Them. Retrieved February 17, 2025, from https://www.kff.org/racial-equity-and-health-policy/issue-brief/racial-disparities-in-maternal-and-infant-health-current-status-and-efforts-to-address-them/

Hirai, A. H., Owens, P. L., Reid, L. D., Vladutiu, C. J., & Main, E. K. (2022). Trends in Severe Maternal Morbidity in the US Across the Transition to ICD-10-CM/PCS From 2012-2019. JAMA Network Open, 5(7), e2222966. https://doi.org/10.1001/jamanetworkopen.2022.22966

Hoogendoorn, S., Oosterbeek, H., & van Praag, M. (2013). The Impact of Gender Diversity on the Performance of Business Teams: Evidence from a Field Experiment [Publisher: INFORMS: Institute for Operations Research]. Management Science, 59(7), 1514–1528. https://doi.org/10.1287/mnsc.1120.1674

Houchens, N., Quinn, M., Harrod, M., Cronin, D. T., Hartley, S., & Saint, S. (2020). Strategies of Female Teaching Attending Physicians to Navigate Gender-Based Challenges: An Exploratory Qualitative Study. Journal of Hospital Medicine, 15(8), 454–460. https://doi.org/10.12788/jhm.3471

Howell, E. A., & Zeitlin, J. (2017). Quality of Care and Disparities in Obstetrics [Publisher: Elsevier]. Obstetrics and Gynecology Clinics, 44(1), 13–25. https://doi.org/10.1016/j.ogc.2016.10.002

Ivanova-Stenzel, R., & Kübler, D. F. (2005). Courtesy and Idleness: Gender Differences in Team Work and Team Competition. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.825686

Johnson, E. M., & Rehavi, M. M. (2016). Physicians Treating Physicians: Information and Incentives in Childbirth. American Economic Journal: Economic Policy, 8(1), 115–141. https://doi.org/10.1257/pol.20140160

Karpowitz, C., O'Connell, S. D., Preece, J., & Stoddard, O. (2024). Strength in Numbers? Gender Composition, Leadership, and Women's Influence in Teams [Publisher: The University of Chicago Press]. Journal of Political Economy. https://doi.org/10.1086/729578

Kearney, E., Razinskas, S., Weiss, M., & Hoegl, M. (2022). Gender diversity and team performance under time pressure: The role of team withdrawal and information elaboration [Publisher: John Wiley & Sons, Inc.]. Journal of Organizational Behavior (John Wiley & Sons, Inc.), 43(7), 1224–1239. https://doi.org/10.1002/job.2630

Kennedy-Moulton, K., Miller, S., Persson, P., Rossin-Slater, M., Wherry, L., & Aldana, G. (2022, November). Maternal and Infant Health Inequality: New Evidence from Linked Administrative Data. https://doi.org/10.3386/w30693

Kevin. (2020, November). Being an attending: What no one tells you in residency and medical school. Retrieved February 20, 2025, from https://kevinmd.com/2020/11/being-an-attending-what-no-one-tells-you-in-residency-and-medical-school.html

Khalil, A., Samara, A., O'Brien, P., Coutinho, C. M., Quintana, S. M., & Ladhani, S. N. (2023). A call to action: The global failure to effectively tackle maternal mortality rates [Publisher: Elsevier]. The Lancet Global Health, 11(8), e1165–e1167. https://doi.org/10.1016/S2214-109X(23)00247-4

La Forgia, A. (2023). The Impact of Management on Clinical Performance: Evidence from Physician Practice Management Companies [Publisher: INFORMS]. Management Science, 69(8), 4646–4667. https://doi.org/10.1287/mnsc.2022.4571

Lamiraud, K., & Vranceanu, R. (2018). Group gender composition and economic decision-making: Evidence from the Kallystée business game [Publisher: Elsevier]. Journal of Economic Behavior & Organization, 145(100), 294–305. Retrieved March 23, 2025, from https://ideas.repec.org//a/eee/jeborg/v145y2018icp294-305.html

Latura, A., & Weeks, A. C. (2023). Corporate Board Quotas and Gender Equality Policies in the Workplace [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12709]. American Journal of Political Science, 67(3), 606–622. https://doi.org/10.1111/ajps.12709

Li, Y., Koopmann, J., Lanaj, K., & Hollenbeck, J. R. (2022). An integration-and-learning perspective on gender diversity in self-managing teams: The roles of learning goal orientation and shared leadership. The Journal of Applied Psychology, 107(9), 1628–1639. https://doi.org/10.1037/apl0000942

Lu, Y., Naik, N. Y., & Teo, M. (2024). Diverse Hedge Funds. Review of Financial Studies, 37(2), 639–683.

Main, E. K., Abreo, A., McNulty, J., Gilbert, W., McNally, C., Poeltler, D., Lanner-Cusin, K., Fenton, D., Gipps, T., Melsop, K., Greene, N., Gould, J. B., & Kilpatrick, S. (2016). Measuring severe maternal morbidity: Validation of potential measures. American Journal of Obstetrics and Gynecology, 214(5), 643.e1–643.e10. https://doi.org/10.1016/j.ajog.2015.11.004

Main, E. K., Morton, C. H., Melsop, K., Hopkins, D., Giuliani, G., & Gould, J. B. (2011). Cesarean Deliveries, Outcomes, and Opportunities for Change in California: Toward a Public Agenda for Maternity Care Safety and Quality. CMQCC.

Matsa, D. A., & Miller, A. R. (2013). A Female Style in Corporate Leadership? Evidence from Quotas. American Economic Journal: Applied Economics, 5(3), 136–169. https://doi.org/10.1257/app.5.3.136

Mazure, C. M. (2021). What Do We Mean By Sex and Gender? Retrieved June 10, 2024, from https://medicine.yale.edu/news-article/what-do-we-mean-by-sex-and-gender/

Mendelberg, T., Karpowitz, C. F., & Goedert, N. (2014). Does Descriptive Representation Facilitate Women's Distinctive Voice? How Gender Composition and Decision Rules Affect Deliberation [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12077]. American Journal of Political Science, 58(2), 291–306. https://doi.org/10.1111/ajps.12077

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. Journal of Personality and Social Psychology, 77(5), 1004–1010. https://doi.org/10.1037//0022-3514.77.5.1004

Ruiz, M. (2014, April). 20 Things No One Tells You About Giving Birth [Section: Lifestyle]. Retrieved May 17, 2025, from https://www.cosmopolitan.com/lifestyle/advice/a6460/things-no-one-told-you-about-giving-birth/

Silver, D. (2021). Haste or Waste? Peer Pressure and Productivity in the Emergency Department. The Review of Economic Studies, 88(3), 1385–1417. https://doi.org/10.1093/restud/rdaa054

Singh, M. (2021). Heuristics in the delivery room [Publisher: American Association for the Advancement of Science]. Science, 374(6565), 324–329. https://doi.org/10.1126/science.abc9818

Singh, M., & Venkataramani, A. (2022, August). Rationing by Race. https://doi.org/10.3386/w30380

Snowden, J. M., Lyndon, A., Kan, P., El Ayadi, A., Main, E., & Carmichael, S. L. (2021). Severe Maternal Morbidity: A Comparison of Definitions and Data Sources. American Journal of Epidemiology, 190(9), 1890–1897. https://doi.org/10.1093/aje/kwab077

Souza, J., Gülmezoglu, A., Lumbiganon, P., Laopaiboon, M., Carroli, G., Fawole, B., Ruyan, P., & the WHO Global Survey on Maternal and Perinatal Health Research Group. (2010). Caesarean section without medical indications is associated with an increased risk of adverse short-term maternal outcomes: The 2004-2008 WHO Global Survey on Maternal and Perinatal Health. BMC Medicine, 8(1), 71. https://doi.org/10.1186/1741-7015-8-71

The Office of Data Dissemination and Transparency. (n.d.). Hospital Discharge Data Dictionary. Retrieved May 15, 2024, from https://quality.healthfinder.fl.gov/Researchers/Order-Data/

Thomas-Hunt, M. C., & Phillips, K. W. (2004). When what you know is not enough: Expertise and gender dynamics in task groups. Personality & Social Psychology Bulletin, 30(12), 1585–1598. https://doi.org/10.1177/0146167204271186

Tonei, V. (2019). Mother's mental health after childbirth: Does the delivery method matter? Journal of Health Economics, 63, 182–196. https://doi.org/10.1016/j.jhealeco.2018.11.006

Tsugawa, Y., Jena, A. B., Figueroa, J. F., Orav, E. J., Blumenthal, D. M., & Jha, A. K. (2017). Comparison of Hospital Mortality and Readmission Rates for Medicare Patients Treated by Male vs Female Physicians. JAMA internal medicine, 177(2), 206–213. https://doi.org/10.1001/jamainternmed.2016.7875

UT Southwestern Medical Center. (2021, March). What if my Ob/Gyn isn't on call to deliver my baby? — Your Pregnancy Matters — UT Southwestern Medical Center. Retrieved June 17, 2025, from http://utswmed.org/medblog/childbirth-my-obgyn-delivery/

White VanGompel, E., Perez, S., Datta, A., Wang, C., Cape, V., & Main, E. (2019). Cesarean overuse and the culture of care. Health Services Research, 54(2), 417–424. https://doi.org/10.1111/1475-6773.13123

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a Collective Intelligence Factor in the Performance of Human Groups [Publisher: American Association for the Advancement of Science]. Science, 330(6004), 686–688. https://doi.org/10.1126/science.1193147

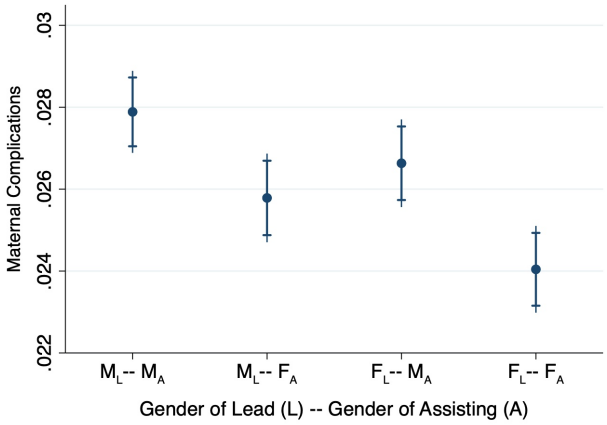
Yang, Y., Tian, T. Y., Woodruff, T. K., Jones, B. F., & Uzzi, B. (2022). Gender-diverse teams produce more novel and higher-impact scientific ideas [Publisher: Proceedings of the National Academy of Sciences]. Proceedings of the National Academy of Sciences, 119(36), e2200841119. https://doi.org/10.1073/pnas.2200841119

Yu, R., Kelz, R., Lorch, S., & Keele, L. J. (2023). The risk of maternal complications after cesarean delivery: Near-far matching for instrumental variables study designs with large observational datasets. Annals of Applied Statistics, 17(2), 1701–1721. https://doi.org/10.1214/22-AOAS1691

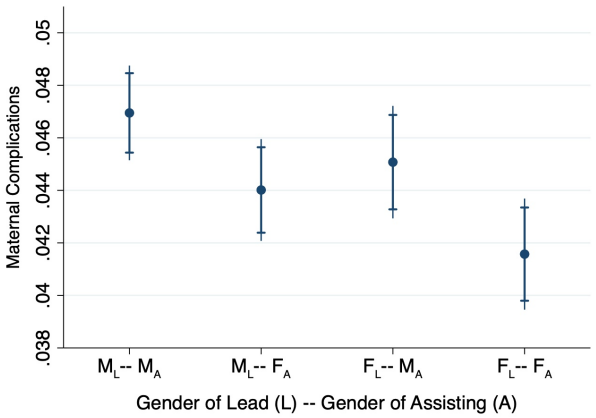
Zureich, J., & Singh, M. (2024, March). Do Physicians Improve More from Positive or Negative Feedback? (Forthcoming in Management Science). https://doi.org/10.2139/ssrn.4746054

FIGURE I
Gender Mix of Physician Teams and Maternal Complications

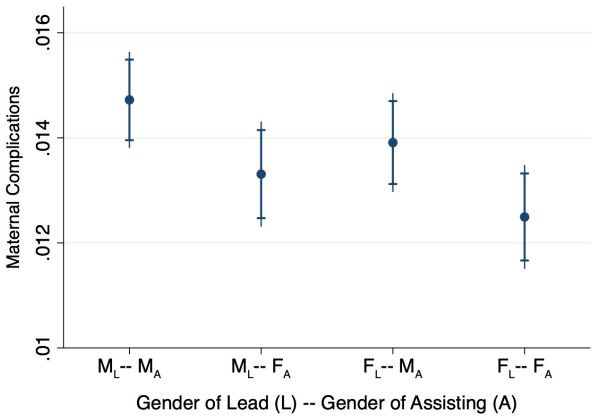
(a) All births



(b) Only C-section births



(c) Only Vaginal births



Notes: *M/F* indicates physician gender, and subscript *L/A* denotes whether the physician is the Lead or Assisting. *Maternal Complications* (MC) is an indicator for the patient experiencing any of 25 adverse events during labor and delivery. Error bars depict 90% and 95% confidence intervals.

Subfigure (a) presents fitted values ($\widehat{MC}$) from Equation 1 for each team type, evaluated at the mean of all covariates (see Column 1 in Table 3 for regression output).

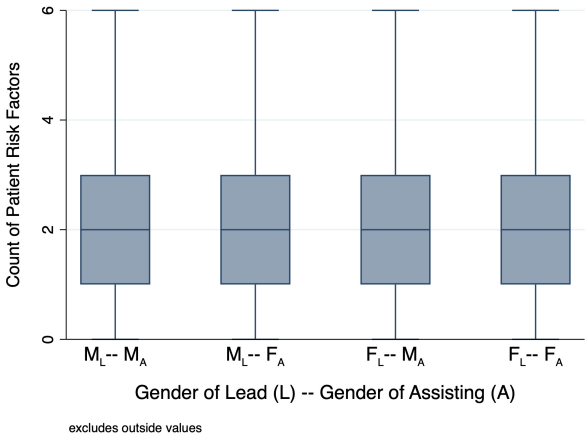
Subfigure (b) replicates Subfigure (a) but restricts the sample to births delivered via Cesarean section.

Subfigure (c) replicates Subfigure (a) but restricts the sample to vaginal deliveries.

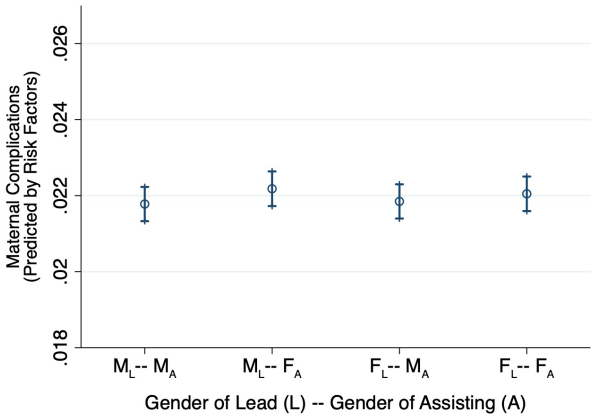
FIGURE II
Support for Identifying Assumption

Panel A. No differences in patient risk

(a) Number of patient risk factors

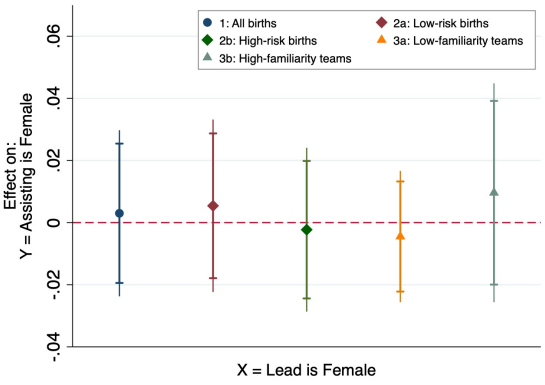


(b) Predicted maternal complications

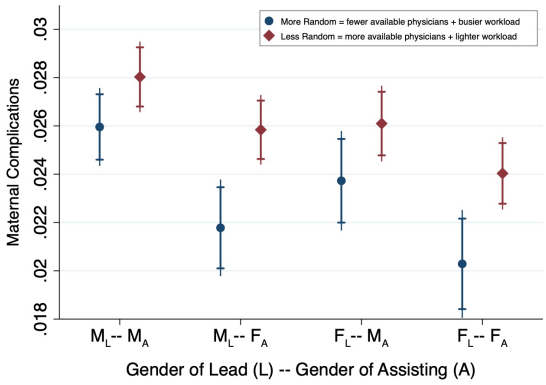


Panel B. No endogenous team formation

(c) By physician gender



(d) By randomness in hospital conditions



Notes: *M/F* indicates physician gender; subscript *L/A* denotes Lead or Assisting. *Maternal Complications* (MC) equals 1 if the patient experiences any of 25 adverse events during delivery. Error bars depict 90% and 95% confidence intervals.

Subfigure (a) Boxplot of patient risk factor counts by team type.

Subfigure (b) Mean predicted maternal complications ($\widehat{MC^{RF}}$) based solely on risk factors, by team type, at sample mean number of recorded diagnoses.

Subfigure (c) Extension of Eq. 1, with the Assisting physician's gender (1 if Female) as the outcome and the Lead physician's gender (1 if Female) as the explanatory variable. Marginal effects shown for:

(1) All births;

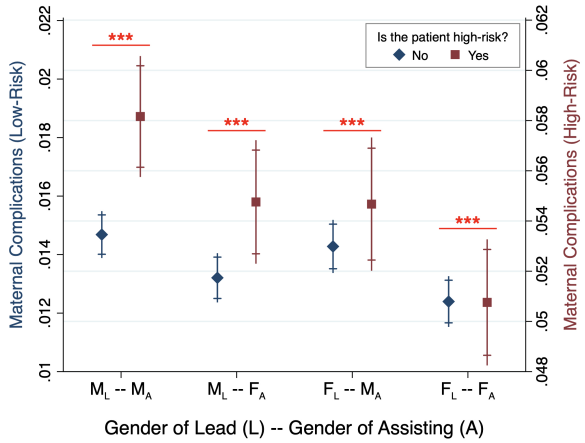
(2) Lead gender interacted with indicator for high-risk patients (above median risk factors);

(3) Lead gender interacted with indicator for high-familiarity teams (above median deliveries together).

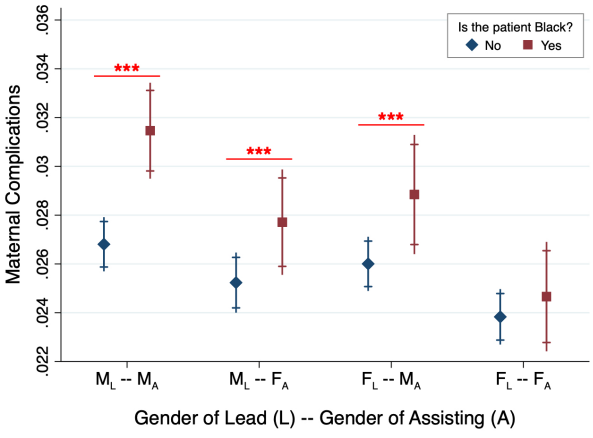
Subfigure (d) Fitted values from extending Equation 1, estimated separately for hospitals with more random team formation (below-median physician count, above-median patient-to-provider ratio) versus less random (above-median physician count, below-median patient-to-provider ratio). See Appendix Section E for details.

FIGURE III
Heterogeneities in Performance

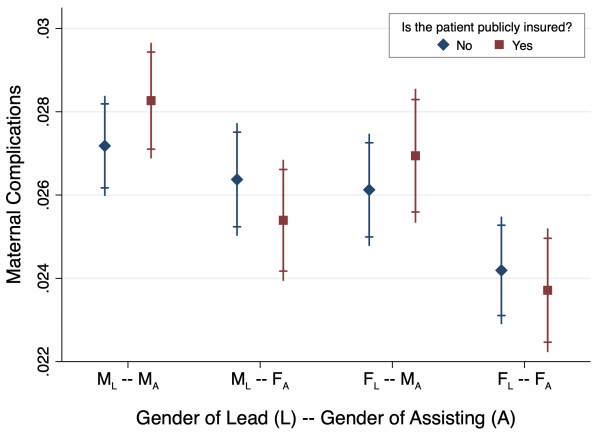
(a) By patient risk



(b) By patient race



(c) By patient insurance



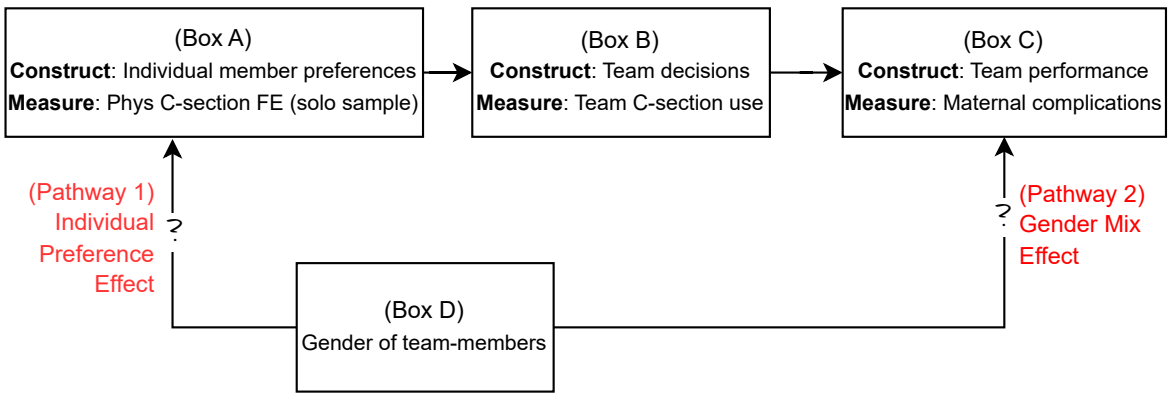
Notes: *M/F* indicates physician gender, and subscript *L/A* denotes whether the physician is the Lead or Assisting. *Maternal Complications* (MC) is an indicator for the patient experiencing any of 25 adverse events during labor and delivery. Error bars depict 90% and 95% confidence intervals. Significance levels: $p < .01^{***}, p < .05^{**}, p < 0.1^{*}$.

Subfigure (a) presents subgroup analyses by patient risk, estimating Equation 1 separately for low- and high-risk patients, defined by whether their number of clinical risk factors is above or below the sample median. Patient controls relevant to maternal complications are included separately within each subgroup. Note the differing y-axis scales, as low-risk patients have substantially lower complication rates.

Subfigure (b) shows differences in maternal complications between Black and non-Black patients using an extended version of Equation 1 with a triple interaction for patient race.

Subfigure (c) replicates Subfigure (b), replacing the racial indicator with an indicator for public (versus private) insurance status.

FIGURE IV
Conceptual Framework: How Does Gender Affect Team Performance?

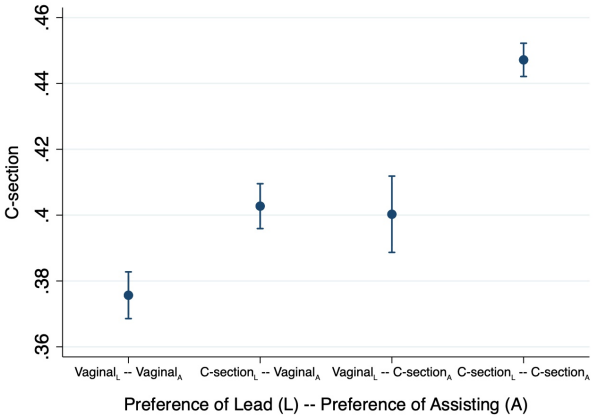


Notes: This figure illustrates our conceptual framework for distinguishing between the Individual Preference effect (Pathway 1) and the Gender Mix effect (Pathway 2). The boxes represent constructs in this framework and their corresponding empirical measures. The Individual Preference effect refers to how gendered differences in individual preferences may affect team decisions and performance. The Gender Mix effect refers to how the gender mix of the team may directly influence team performance independently of individual physician preferences.

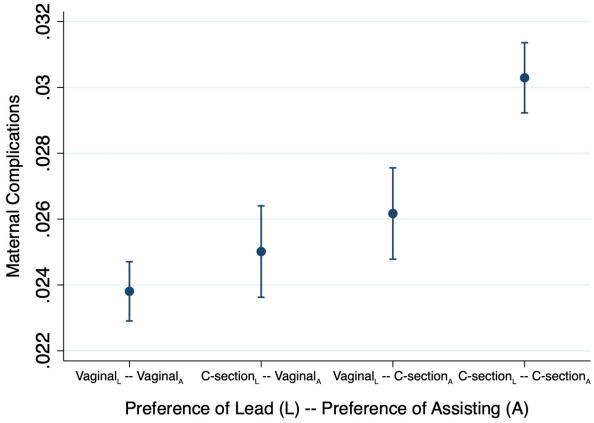
FIGURE V
The "How:" Exploring Mechanisms via C-section Decisions

Panel A. The Effect of Physician Preferences on:

(a) Team decision: C-section

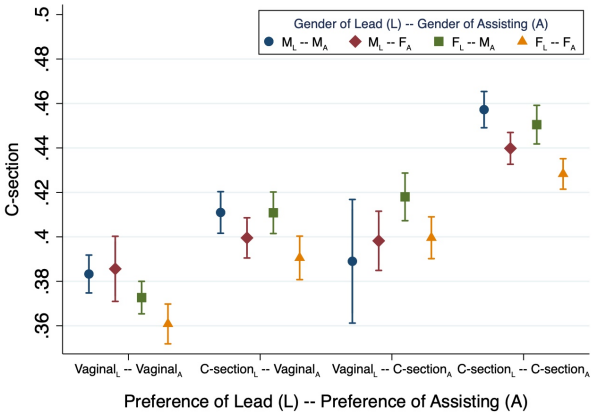


(b) Team performance: Maternal complications

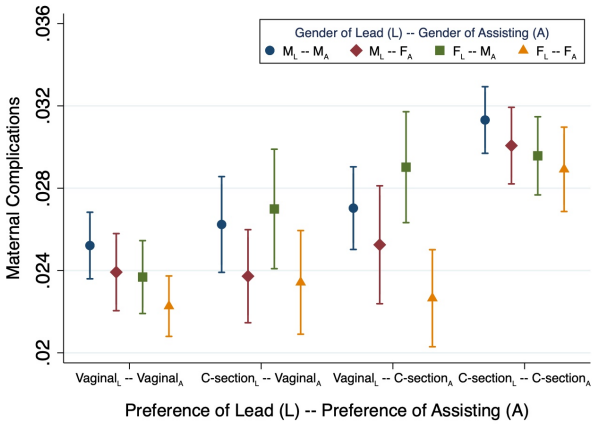


Panel B. The Effect of Gender Mix of the Team on:

(c) Team decision: C-section



(d) Team performance: Maternal complications



Notes: *M/F* indicates physician gender, *Vaginal/C-section* indicates the physician's preference either for vaginal delivery or C-section based on the solo sample of births, and subscript *L/A* denotes whether the physician is the Lead or Assisting. Error bars depict 90% and 95% confidence intervals.

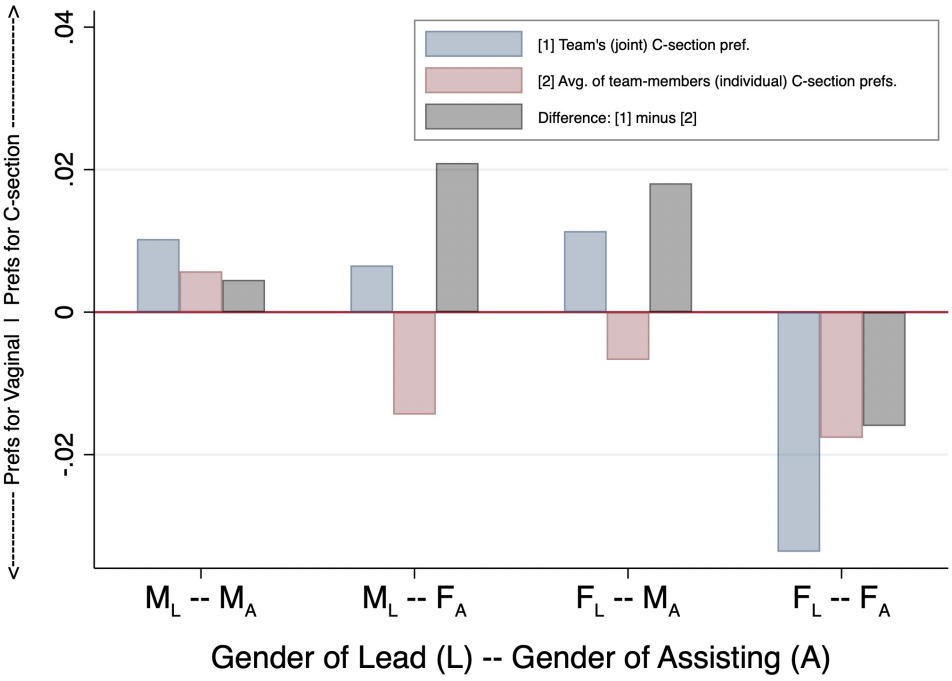
Panel A presents fitted values from Equation 2 for each combination of physician preferences (each team-member has a preference for either C-section or Vaginal delivery).

Panel B recreates Panel A, where each estimate in Panel A is now disaggregated by team type.

Subfigures (a) and (c) use *C-section* as the outcome, which equals 1 if delivery was via C-section and 0 if vaginal.

Subfigures (b) and (d) use *Maternal Complications* as the outcome, defined as an indicator equal to 1 if the patient experienced any of 25 adverse events during labor and delivery.

FIGURE VI
The "Why": Gender Mix Affects Team Incorporation of Individual Preferences

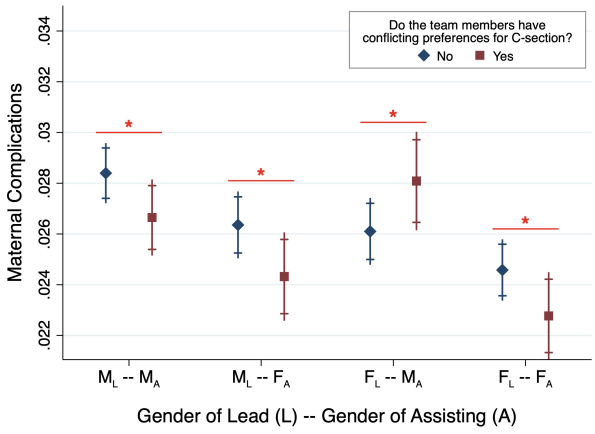


Notes: M/F indicates physician gender, and subscript L/A denotes whether the physician is the Lead or Assisting. Positive (vs. negative) values on the Y axis signify that the individuals or team prefers C-section (vs. vaginal delivery).

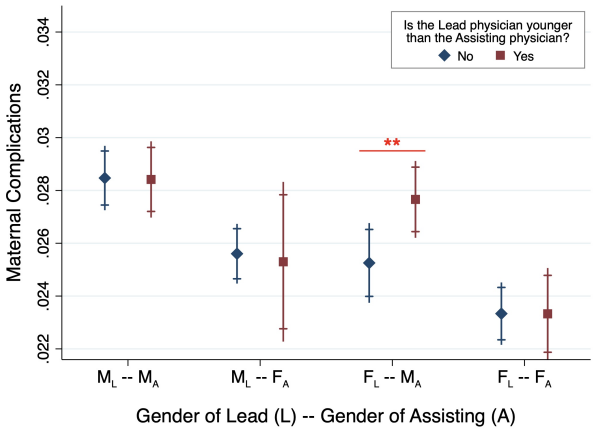
For each team type, this figure shows:
(1) The team's joint C-section preference, measured by the team fixed effect obtained from the team sample.
(2) The average of the Lead and Assisting physicians' individual preferences for C-sections, measured by individual physician fixed effects from the solo sample.
(3) The difference between (1) and (2).

FIGURE VII
The "Why": Gender Mix Affects Team Resilience to Challenges

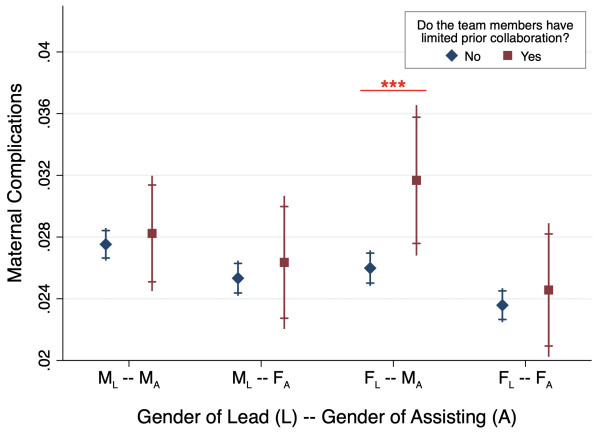
(a) Challenge: Conflicting preferences



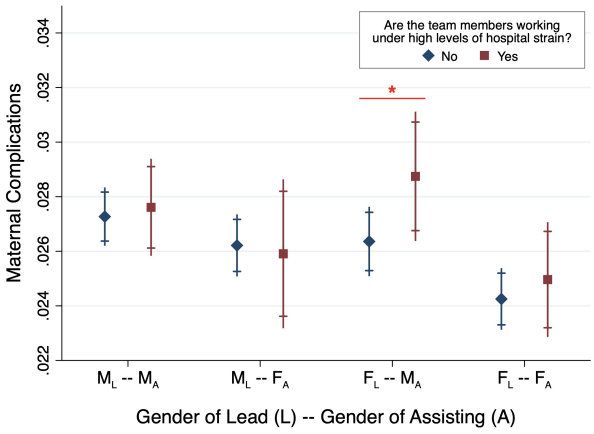
(b) Challenge: Lead younger than Assisting



(c) Challenge: Limited prior collaboration



(d) Challenge: High hospital strain



Notes: M/F indicates physician gender; subscript L/A indicates Lead or Assisting physician. Error bars depict 90% and 95% confidence intervals. Significance levels: $p < .01^{***}, p < .05^{**}, p < 0.1^{*}$.

Fitted values of maternal complications ($\widehat{MC}$) from Appendix Equation A.6 are presented at covariate means. Each subfigure uses a different indicator variable (Var) if:

(a) Physicians have conflicting delivery preferences (one prefers vaginal birth, one prefers C-section).
(b) Assisting physician is at least 5 years older than Lead.
(c) Team has delivered 2 babies together within first two recorded quarters (5th percentile of volume).
(d) Hospital birth volume in top quintile for the given quarter-year.

TABLE 1
Summary Statistics: Patients

	Full Sample	Solo Sample	Team Sample
Outcome Variables (Raw Mean)			
Maternal Complications	2.07%	1.91%	2.61%
C-Sections	40.68%	40.75%	40.47%
Patient Characteristics			
Privately insured	42.73%	42.07%	44.95%
Medicaid	49.57%	50.06%	47.90%
Other insurance	7.70%	7.87%	7.15%
Black	21.56%	21.32%	22.37%
Hispanic/Latina	24.56%	24.09%	26.15%
White	46.11%	46.93%	43.34%
Other race	7.77%	7.65%	8.15%
Patient Risk Factors			
Patient Age			
Age ($\leq$22)	20.44%	20.50%	20.24%
Age (23-26)	21.02%	21.14%	20.59%
Age (27-30)	23.12%	23.05%	23.34%
Age (31-33)	19.24%	19.14%	19.56%
Age ($\geq$34)	16.19%	16.17%	16.27%
Asthma	3.27%	3.07%	3.96%
Anemia	12.21%	11.57%	14.40%
Poly- & Oligo- hydramnios	4.08%	3.78%	5.09%
Maternal physical abnormality	7.24%	6.82%	8.65%
Blood disorders	3.29%	3.09%	4.00%
Uterine size issue	0.28%	0.23%	0.45%
Infant size issue	6.02%	5.89%	6.44%
Obesity	4.68%	4.31%	5.94%
Diabetes	7.08%	6.70%	8.34%
Substance abuse or smoking	9.01%	8.90%	9.37%
Infectious and parasitic conditions	4.07%	3.93%	4.55%
Heart disease	0.88%	0.81%	1.10%
Known fetal abnormality	1.87%	1.68%	2.54%
Hypertension	11.15%	9.96%	15.19%
Isoimmunization	1.98%	1.95%	2.09%
Other conditions/risks	2.60%	2.42%	3.21%
Ruptured membrane	4.41%	3.86%	6.28%
Previous pregnancy	24.73%	24.56%	25.28%
Breech	8.15%	7.95%	8.84%
Multiple gestation	1.84%	1.76%	2.09%
Pre-term birth	6.70%	6.15%	8.54%
Previous C-section	19.66%	21.22%	14.35%
Observations	2507736	1936754	570982

Notes: Summary statistics of patient characteristics and risk factors for the full sample, solo sample (births with one physician), and team sample (births with two physicians). See Section C for detailed variable definitions.

TABLE 2
Summary of Paper: All Analyses and Findings

	Results and relevant Tables/ Figures
Main analysis	
Does team gender mix affect performance (maternal complications)?	Figure I: Gender mix affects team performance. Being in a same-gender team improves performance for women, but harms performance for men. Maternal complication rates from lowest to highest are: F_L–F_A < M_L–F_A < F_L–M_A < M_L–M_A. We focus on two main findings: 1. Among same-gender teams, F_L–F_A perform better than M_L–M_A. 2. Among mixed-gender teams, F_L–M_A perform worse than M_L–F_A.
Robustness checks	
Are these results robust?	Results are robust to: sparse and saturated controls (Table 3); logistic regression (Table A.2); varying analytic samples (Table A.3); varying outcome definitions (Table A.4); and including team or physician fixed effects (Figure A.2).
Support for identifying assumption	
Can the effect of team gender mix on maternal complications be interpreted causally? Identification assumption: Quasi-random assignment to *gender mix* of team	We provide 3 types of evidence in support of identification: 1. Conceptual support: There is randomness in timing of birth and pre-set physician schedules. 2. There are no differences in patient risk across the 4 team types (Figure II Subfigures (a) and (b), and Figure A.5). 3. There is no evidence of endogenous team formation by physician gender, patient risk, team familiarity, or physician skill. (Figure II Subfigures (c) and (d), and Figure A.6).
Heterogeneities in performance	
Are there heterogeneities in performance by patient (i) risk; (ii) race; (iii) insurance?	1. Results hold for low- and high-risk patients (Figure III Subfigure (a)). 2. Results hold for Black and non-Black patients, but only F_L–F_A teams achieve the lowest maternal complication rates for, and show no disparity between, Black and non-Black patients (Figure III Subfigure (b)). 3. Results hold for private- and Medicaid-insured patients (Figure III Subfigure (c)).
Mechanisms: How does gender mix affect team performance?	
Gender mix of the team can affect performance through two pathways: (1) The *Individual Preference* effect: Male and female physicians have different preferences for clinical decisions, so team decisions simply reflect the mix of preferences. (2) The *Gender Mix* effect: The genders of the team-members directly and causally affect team decisions and performance, holding constant the preferences of the team members. Can we determine which of these two pathways explains our main results?	Focusing on the discretionary decision to perform a C-section, we estimate each physician's preference for C-sections by isolating their C-section fixed effect on their sample of solo deliveries (Figure A.7). We find that, while individual physician C-section preferences are highly predictive of team decisions and outcomes, the team's gender mix directly affects the C-section decision and maternal complication rates *even holding constant each physician's preference.* That is, the Gender Mix effect is driving our main results (Figure V).
Mechanisms: Why does gender mix affect team performance?	
Gender mix of the team may introduce *gender norms* that affect team decision-making and performance. Are there patterns in the data that are consistent with this explanation?	(1) Gender mix affects the incorporation of individual preferences into team decisions. $\Longrightarrow$ Figure VI: F_L–F_A teams' superior performance may be explained by team decisions closely reflecting average member preferences, and performing *fewer* C-sections as a team than they would individually. This may reflect gender norms, such as men's higher risk tolerance/action-orientation and undervaluation of women's opinions in groups. (2) Gender mix affects team resilience to challenging conditions. $\Longrightarrow$ Figure VII: F_L–M_A teams are the least resilient, exhibiting higher maternal complication rates when team members have conflicting preferences, the Lead physician is younger than the Assisting, prior collaboration is limited, and hospital strain is high. This may result from F_L–M_A teams inverting traditional gender norms around leadership, potentially creating baseline teamwork frictions.

TABLE 3
Gender Mix of Physician Teams and Maternal Complications

	Main Specification	Robustness				
	(1)	(2)	(3)	(4)	(5)	(6)
Female Lead	0.0008	-0.0009	0.0004	0.0006	0.0008	0.0014
	(0.0008)	(0.0010)	(0.0008)	(0.0008)	(0.0008)	(0.0014)
Same Gender Assisting	0.0021***	0.0006	0.0015**	0.0012	0.0020**	0.0024*
	(0.0008)	(0.0008)	(0.0007)	(0.0008)	(0.0008)	(0.0013)
Female Lead × Same Gender Assisting	-0.0047***	-0.0032***	-0.0047***	-0.0036***	-0.0044***	-0.0042**
	(0.0011)	(0.0011)	(0.0010)	(0.0011)	(0.0011)	(0.0019)
Physician Controls	Yes	No	No	Yes	Yes	Yes
Patient Controls	Yes	No	Yes	Yes	Yes	Yes
Patient Zip Fixed Effects	No	No	No	No	Yes	Yes
Year Fixed Effects	No	Yes	Yes	Yes	No	No
Hospital Fixed Effects	No	Yes	Yes	Yes	No	No
Quarter×Year Fixed Effects	Yes	No	No	No	Yes	No
Hospital×Year Fixed Effects	Yes	No	No	No	Yes	No
Hospital×Quarter×Year×DayofWeek×AdmHour FE	No	No	No	No	No	Yes
Ymean	0.026	0.026	0.026	0.026	0.026	0.020
Observations	540400	570982	570982	540411	539021	273646
R^2	0.093	0.007	0.063	0.089	0.096	0.375

Notes: Regression coefficients and standard errors (in parentheses) from Equation 1, with sequential inclusion of patient controls, physician controls, and various fixed effects. The outcome variable, *Maternal Complications*, equals 1 if the patient experiences any of 25 adverse events during labor and delivery. *Female Lead* equals 1 if the Lead physician is female, 0 otherwise. *Same Gender Assisting* equals 1 if the Assisting physician shares the Lead's gender, 0 otherwise. *DayofWeek* refers to the day the patient was admitted to the hospital and *AdmHour* is the hour of the day the patient was admitted to the hospital. Significance levels: $p < .01^{***}, p < .05^{**}, p < 0.1^{*}$.

Online Appendix

A Data and Sample

The primary data used in this research was inpatient discharge records from the Florida Agency for Healthcare Administration (AHCA). This data can be accessed via a Data Use Agreement with AHCA (apply here: https://quality.healthfinder.fl.gov/Researchers/Order-Data/). These data include all patients that delivered a baby in a Florida hospital between 2006 and 2018. Florida also provides publicly available data (referred to as the Florida Licensure data) on physician characteristics, including physician date of birth, name, specialty, medical license number, medical school, and graduation dates (download here: https://flhealthsource.gov/data-portal/). Physicians must respond to these Florida surveys every two years in order to maintain an active license. We obtained SK&A data from 2006-2016 via Wharton WRDS. The SK&A includes information on physician gender, date of birth, location, and specialty for all office-based physicians in the United States. SK&A has a research center that verifies physician information through telephone surveys every six months, but the information is also gathered through physician websites, state licensing information, mergers and acquisitions data, professional associations, and government agencies. We downloaded Medicare Physician Compare data for 2014-2018 (2014 is the first year the data is available; download here: https://data.cms.gov/provider-data/archived-data/doctors-clinicians). For physicians who bill Medicare (which not all OB-GYNs do), this dataset provides physician name, gender, specialty, and location information. Therefore, using these three datasets, we are able to determine age, gender[17], and specialty information for nearly all OB-GYNs in our sample.

Figure A.1 demonstrates the sample restrictions, which take us from 2,788,011 births from 2006-2018 in the full sample to 570,982 births delivered by two physicians. Births were determined as follows: DRGs 370-375 (2000 - 2007Q123); MSDRGs 765 -768, 774, 775 (2007Q4 - 2018Q123); 2018Q4 MSDRGs (768, 783-788, 796-798, 805-807). Table A.1 provides the characteristics of physician teams in the sample. Note that we use physician and "OB-GYN" interchangeably. There are 1,010 female physicians and 1,034 male physicians (2,044 total physicians). We imputed gender from first names for 4.0% of physicians using data from the US Social Security Administration and the Gender Checker Directory. Similarly, for the 12.8% of physicians with missing age data, we manually collected their age from online physician profiles on www.sharecare.com, a website that uses American Medical Association data to determine the date of birth. For the remaining 2.3% of physicians for whom we could not find an age, we imputed their age using their medical school graduation date by assuming they were 28 years old at graduation. Ultimately, slight mismeasurement in age would not influence our findings because we create quintiles of physician age to include in regression analyses.

Table 1 shows the summary statistics for patient characteristics in the full sample, solo sample, and team sample. Overall, the different samples show a similar distribution of patients in terms of insurance status, race/ethnicity, and patient age. However, patients in the team sample exhibit higher rates of maternal complications, consistent with teams likely attending more complicated births than when a physician is delivering solo.

[17] Following the recommendation of the Institute for Medicine, we use the term gender "to refer to a person's self-representation as male or female" (Mazure, 2021). Our gender data comes from physician survey responses, capturing individuals' self-reported gender and not sex at birth. However, we acknowledge that different disciplines may prefer different languages.

B Physician Roles

In the AHCA data, physician identifiers are available for the "attending" physician, the "operating or performing" physician, and "other" physician listed on the patient record (we drop births with a unique third physician, see our sample selection in Figure A.1). We refer to the attending as the "Lead" and the operating as the "Assisting" based on the legal definitions and conversations with AHCA to confirm the physician roles. The attending (i.e., the Lead physician) is defined as the clinician "who had primary responsibility for the patient's medical care and treatment or who certified as to the medical necessity of the services rendered." The attending physician is typically the main OB-GYN on call on a labor and delivery ward. The operating or performing physician (i.e., the Assisting physician) is defined as the clinician "who had primary responsibility for the principal procedure performed." The operating or performing physician may be the attending physician. For a vaginal delivery involving a team, the most common principal procedure performed is "Other manually assisted delivery" (ICD9 code 73.59), which then became "Delivery of Products of Conception, External Approach" (ICD 10 code 10E0XZZ). Other common procedures include delivery with forceps or vacuum, or episiotomy. For a C-section delivery involving a team, the most common principal procedure is "a low cervical cesarean section" (ICD9 code 74.1), which then became "Extraction of Products of Conception, Low, Open Approach" (ICD10 code 10D00Z1). Therefore, the Assisting physician is directly involved in delivering the baby in both vaginal births and C-sections.

While more colloquially the term "attending" is used to describe a physician supervising fellows, residents, and medical students, this is not the case in inpatient discharge records since every single inpatient encounter in all hospitals has an attending physician recorded (i.e., even in non-teaching hospitals, the primary physician on the patient record is referred to as the attending physician). Importantly, the attending physician can also have operated or performed a procedure. To avoid confusion surrounding this language, and to better match the definitions of the data, we therefore refer to the attending as the "Lead" physician and the operating as the "Assisting" physician.

C Variable Construction

C.1 Maternal complications.

We code maternal complications following the definition of "severe maternal morbidity" (SMM) developed by the Centers for Disease Control and Prevention International Classification of Diseases, 9th revision (Main et al., 2016). Specifically, we create a binary indicator for whether a patient experiences a maternal complications if they experience any of the following 25 conditions: Acute myocardial infarction, acute renal failure, adult respiratory distress syndrome, amniotic fluid embolism, aneurysm, cardiac arrest/ventricular fibrillation, disseminated intravascular coagulation, eclampsia, heart failure during procedure or surgery, internal injuries of thorax, abdomen, and pelvis, intracranial injuries, puerperal cerebrovascular disorders, pulmonary edema, severe anesthesia complications, sepsis, shock, sickle cell anemia with crisis, thrombotic embolism, blood transfusion, cardio monitoring, conversion of cardiac rhythm, hysterectomy, operations on heart and pericardium, temporary tracheostomy, ventilation.

We use the 25-condition version of SMM because most births in our sample (2006-2015Q3) occur during the time period when that was the main indicator. To map from ICD-9 to ICD-10 codes we use crosswalks from the NBER (https://www.nber.org/research/data/

icd-9-cm-and-icd-10-cm-and-icd-10-pcs-crosswalk-or-general-equivalence-mappings) to account for this change and follow CMS guidance for optimizing the mapping (https://www.cms.gov/medicare/coding-billing/icd-10-codes/2018-icd-10-cm-gem). This bi-directional mapping has been used by other researchers to track SMM across the ICD-9 to ICD-10 switch (Hirai et al., 2022).

However, there are several additional measures for SMM that have been discussed in the clinical literature (Snowden et al., 2021). After the transition to ICD-10 codes, the 25 conditions were reshaped into 21 conditions (some conditions were combined, but some codes also changed) (CDC, 2024). While many OB-GYNs support this change, others believe this definition (and even the original definition) to be too restrictive. Therefore, adapting the approach of Snowden et al. (2021), in Table A.4 we show robustness to a more comprehensive definition which includes additional instances of maternal complications developed by Bateman et al. (2013), the more comprehensive definition excluding blood transfusions, and we present results for our main measure of maternal complications using only data from before the change to ICD-10 codes in October 2015.

C.2 Classification of delivery method.

The hospital discharge records associated with each delivery identify whether the delivery was vaginal or C-section. ICD-9-CM Delivery MS-DRG/DRG codes 370, 371, 765, and 766 identify C-sections, and codes 372, 373, 374, 375, 767, 768, 774, and 775 identify vaginal births. To classify whether a birth was planned or unplanned, we follow the methods of Henry et al. (1995) and Gregory et al. (2002). This methodology uses diagnosis codes that indicate a trial of labor and defines a planned C-section as a C-section with no indication of a trial of labor. For example, if a women is recorded as having "failed to progress to labor" then she was allowed to have labored before a C-section, so it was not scheduled in advance.

C.3 Classification of maternal risk factors

In regression analysis, we control for 23 maternal risk factors observable to the physician before the onset of labor. These risk factors have been used in previous research on childbirth outcomes (Currie & MacLeod, 2017; Gregory et al., 2002; Henry et al., 1995; Johnson & Rehavi, 2016; La Forgia, 2023). These 23 risk factors include: Patient age (divided into quintiles of age), asthma, anemia, polyhydramnios or oligohydramnios, maternal physical abnormalities (includes thyroid abnormality, bone or joint disorder, or abnormality of organs and soft tissues of pelvis), blood disorders (includes antepartum hemorrhage, abruptio placentae, placenta previa, uterine rupture, coagulation defects complicating pregnancy and spotting complicating pregnancy), uterine size issue (includes uterine size date discrepancy and cervical shortening), infant size issue (i.e., fetal growth issue affecting the management of the mother), obesity, diabetes or abnormal glucose tolerance, substance abuse or smoking, infectious and parasitic conditions, maternal congenital and other heart disease, known or suspected fetal abnormalities affecting the management of the mother, hypertension complicating pregnancy (includes pre-eclampsia), isoimmunization, premature rupture of membrane or amniotic cavity infection, previous pregnancy, malposition or malpresentation of fetus (includes breech birth), multiple gestation (i.e., twins or above), pre-term birth, previous C-section and other conditions and risks (includes excessive weight gain during pregnancy, habitual aborter, renal disease, liver disorders, nerve disorders, and severe urinary tract infection.

In ICD 9/10 codes, there is no uniform method to account for previous vaginal delivery. In-

stead, "previous pregnancy" captures any instance of previous pregnancy not resulting in a C-section, including any indication the patient had a previous birth (multigravida or grand multiparity), had a previous ectopic pregnancy, or had a pregnancy resulting in stillbirth.

D Empirical Specification

D.1 Main specification

Here we provide more details on our main specification:

$$y_{ijkht} = \beta_1 \text{lead_fem}_j + \beta_2 \text{samegender_assist}_k + \beta_3 (\text{lead_fem}_j \times \text{samegender_assist}_k) + \gamma X_i + \delta Z_j + \phi T_k + \alpha_t + \alpha_{hy} + \varepsilon_{ijkht}$$

- β_1 = is the *effect* of having a female Lead when $\text{samegender_assist}_k = 0$, i.e., when the Assisting is a different gender than the Lead. Thus, it can be described as the following difference: Complications of FM minus (-) the complications of MF, conditional on covariates.
- β_2 = is the effect of the Assisting being the same gender as the Lead, when $\text{lead_fem}_j = 0$, i.e., when the Lead is a male. Thus, it can be described as the following difference: Complications of MM minus (-) the complications of MF.
- β_3 = can be described in two equivalent ways:
  - It is difference between [effect of $samegender_assist_k$ when $lead_fem_j = 1$] minus [effect of $samegender_assist_k$ when $lead_fem_j = 0$]. Thus, it can described as the following difference-in-difference: [Complications of FF - complications of FM] - [Complications of MM - complications of MF].
  - It can be equivalently described as the difference between [effect of $lead_fem_j$ when $samegender_assist_k = 1$] minus [effect of $lead_fem_j$ when $samegender_assist_k = 0$]. Thus, it can described as the following difference-in-difference: [Complications of FF - Complications of MM] - [Complications of FM - Complications of MF], which is simply a rearrangement of the equation above.

Another way to think about β_3 is that it captures whether the difference in maternal complications between the two same-gender teams is significantly larger (positive sign) or smaller (negative sign) than the two mixed-gender teams. It is important to note that β_3 is only comparing *two differences*; it does not actually tell us about the absolute performance of any single team. In other words, a negative sign on β_3 does not necessarily mean lower rates of maternal complications for any one group.

D.2 Additional specifications using team or physician fixed effects

We rely on the main regression above to compare outcomes across all four team types simultaneously and use the entire sample of deliveries. However, to more precisely examine differences in outcomes within teams and within physicians, we also provide results using team or physician fixed effects.

First, we restrict the sample to mixed gender teams and include a fixed effect for the two-physician team (regardless of their role in the team) and examine differences when the female vs male physician in the Lead role in the same team. This "within-team" regression is as follows:

$$y_{ijkht} = \beta \text{lead_fem}_j + \gamma X_i + \phi T_{jk} + \alpha_{jk} + \alpha_t + \alpha_{hy} + \varepsilon_{ijkht} \tag{A.1}$$

where y is the observed outcome for patient i under the care of Lead physician j and Assisting physician k in hospital h and quarter-year t. The β captures the performance of the same mixed-gender team when a female physician is in the Lead role and a male physician is in the Assisting role compared to when the same male physician is in the Lead role and the same female physician is in the Assisting role. As before, X_i is a vector of patient risk factors observable before the onset of labor, including patient age quintiles, risk factors, and socio-demographic characteristics. For the physician pair (T_{jk}), we include a measure of previous collaboration by including quintiles of cumulative births up to the previous quarter performed by the two physicians and include fixed effects for the number of diagnosis codes recorded in the patient's medical record. Lastly, we include quarter-year fixed effects (α_t), hospital-year fixed effects (α_{hy}), and a physician team fixed effects (α_{jk}), regardless of which physician is in the Lead or Assisting physician role. Results of this regression are presented in Figure A.2 Subfigure (a).

Second, we restrict the sample to either Lead female physicians or Lead male physicians and examine differences in outcomes when the Lead is paired with a physician of the same or opposite gender. This "within-physician" regression is as follows:

$$y_{ijkht} = \beta \text{same_gender}_{jk} + \gamma X_i + \phi T_j + \alpha_j + \alpha_t + \alpha_{hy} + \varepsilon_{ijkht} \tag{A.2}$$

where y is the observed outcome for patient i under the care of Lead physician j and Assisting physician k in hospital h and quarter-year t. The β captures the difference in outcomes when the same Lead physician is paired with an Assisting physician of the same gender compared to opposite gender. The regression is estimated separately for Lead female and Lead male physicians. The main difference from above is that we include fixed effects for the Lead physician (α_j). Results of this regression are presented in Figure A.2 Subfigure (b).

E Identification Strategy

To capture the unbiased effect of team gender mix on patient outcomes, we would ideally randomly assign patients to care teams of varying gender mix. Since this is not possible, our identification strategy assumes that the gender mix of the physician team assigned to a patient is plausibly exogenous. In this section, we discuss the many tests and checks we perform in support of this assumption.

E.1 Does patient risk of complication differ by team gender mix?

One concern is that a certain team type is being matched to patients of greater complexity. For example, our identification strategy may be compromised if teams with male OB-GYNs systematically treat patients who are of higher medical complexity and therefore, are more likely to experience a complication. For this reason, we examine whether patients systematically differ across team types by estimating each patient's predicted probability of a maternal complication based *solely* on their observable risk factors and the number of diagnosis codes on their record – what we refer to in the paper as $\widehat{MC^{RF}}$ – using the full sample of births. To do so, we run the following logistic regression, with output presented in Table A.5 for maternal complications:

$$Pr(MC_i^{RF}) = F(\beta^r \mathbf{X}_i^r + \gamma \mathbf{D}_i) \tag{A.3}$$

where $\mathbf{X_i^r}$ includes indicators for all 23 antepartum patient risk factors listed in Section C as well as a

vector of indicator variables ($\mathbf{D}_i$) for the number of diagnosis codes recorded in a patient record (values range from 1 to 31) because coding behavior can vary by physician and influence the number of risk factors reported. That is, the more thorough the coding behavior of the physician, the more likely a risk factor is to be recorded, which may spuriously inflate a patient's $\widehat{MC^{RF}}$. To summarize, $\widehat{MC^{RF}}$ measures an individual patient's probability of experiencing a complication only based on observable clinical risk factors (controlling for the number of diagnosis codes on the record).

In Figure II Subfigure (a) we show that a patient's raw count of risk factors is similar across team types, and in Subfigure (b) we show that the mean $\widehat{MC^{RF}}$ (at the sample mean of the number of diagnosis codes) is also similar for each team type. If we do not account for coding differences across physicians as in Figure II Subfigure (b), we instead find evidence that M_L–M_A teams have lower $\widehat{MC^{RF}}$ and female physicians have higher $\widehat{MC^{RF}}$ (Figure A.4 Subfigure (a)). This suggests that male physicians code less thoroughly than female physicians, which is verified empirically: male-only teams record slightly fewer diagnosis codes on average (5.24) than female-only teams (5.95). This may mechanically reduce the likelihood that any of the 23 patient risk factors we use as covariates are recorded on the patient's record and may thereby make male physicians' patients look less risky than they truly are. However, Figure II Subfigure (b) shows that, if we accounted for differences in the number of codes recorded across patients, all four team types have the same predicted risk of maternal complications based on the 23 observable risk factors.

Not accounting for the coding differences discussed above would have biased *against* our main finding (i.e., biased towards the null) that M_L–M_A teams have higher estimated $\widehat{MC}$ than F_L–F_A teams. In other words, if we do not account for M_L–M_A teams coding less than F_L–F_A teams, we observe that the patients of M_L–M_A teams have the lowest $\widehat{MC^{RF}}$ while patients of F_L–F_A teams have the highest $\widehat{MC^{RF}}$, as shown in Figure A.4 Subfigure (a). If so, our main findings appear even more shocking: we essentially find that M_L–M_A teams have worse outcomes despite treating a lower-risk population, as seen in Figure A.4 Subfigure (b). Fixing the number of diagnosis codes to the sample mean as in Figure A.4 Subfigure (c) and (d), we can see there are no longer any differences in $\widehat{MC^{RF}}$ across team types, but our main finding and pattern remains unchanged: M_L–M_A teams have the highest estimated $\widehat{MC}$ and F_L–F_A teams have the lowest estimated $\widehat{MC}$. The fact that $\widehat{MC^{RF}}$ changes across team types after accounting for coding behavior– but $\widehat{MC}$ does not – makes sense. That is because $\widehat{MC^{RF}}$ is a monotonic and linear function of 23 different antepartum risk factors, while $\widehat{MC}$ is an indicator for *any* of 25 serious clinical conditions, depends on many other factors that are unaffected by coding behavior, and omitting these 25 serious conditions from the coding record would reduce a physician's reimbursement levels. Thus while risk factors like obesity may be omitted from the coding record, serious conditions that comprise the maternal complication measure are unlikely to be omitted. Overall, these figures provide compelling evidence that four team types treat patients with similar risk profiles.

E.2 Are physician teams being formed endogenously?

Team formation by physician gender

Here we provide details on the empirical analyses we use to examine whether Assisting physicians are being assigned to or selected by a Lead physician according to their gender. To test this, we run the following regression:

$$\text{Assisting_fem}_k = \beta \text{Lead_fem}_j + \gamma X_i + \phi T_j + \alpha_t + \alpha_{hy} + \alpha_n + \varepsilon_{ijkht} \tag{A.4}$$

The β captures the difference between a female and male Lead in the probability that the Assisting physician is female. As before, X_i is a vector of patient risk factors observable before the onset of labor, including patient age quintiles, risk factors, and socio-demographic characteristics. For the Lead physician (T_j), we include quintiles of physician age and quintiles of cumulative births up to the previous quarter and fixed effects for the number of diagnosis codes recorded in the patient's medical record. Lastly, we include quarter-year fixed effects (α_t), hospital-year fixed effects (α_{hy}), and fixed effects for the total number of unique male and female assisting physicians in a hospital in a quarter-year (α_n). Fixed effects for the number of male and female physicians in a quarter account for the mechanical likelihood of gender-based matching (i.e., a male Lead will be less likely to be paired with a female Assisting if male physicians outnumber female physicians). However, results are also robust to excluding these fixed effects.

Results of this regression are presented in Figure II Subfigure (c). As an extension of the regression above, we interact $Lead_fem_j$ with a variable for whether the patient is of high or low risk (above-median or below-median number of clinical risk factors) as well as with a variable for whether the Lead and Assisting physician are more or less familiar with each other (above-median or below-median familiarity based on the cumulative number of previous births delivered by the team.)

Team formation by physician skill

Here we provide details on the empirical analyses we use to examine whether physicians of different genders are being matched to teams based on a proxy for their skill. To proxy for skill, we use the physician solo sample (i.e., only one physician is listed on the birth record) and run the following regression using a linear probability model:

$$MC_{ijht} = \gamma X_i + \alpha_j + \alpha_t + \alpha_{hy} + \varepsilon_{ijht} \tag{A.5}$$

where MC is the observed maternal complications for patient i under the care of physician j in hospital h and quarter-year t. X_i is a vector of 23 patient risk factors observable before the onset of labor, (α_t) represents quarter-year fixed effects, (α_{hy}) hospital-year fixed effects, and (α_j) represents physician fixed effects.

The physician fixed effects obtained from this regression can be interpreted as a proxy for physician skill because they represent how much of a patient's risk of maternal complications is explained by time-invariant physician-specific factors after accounting for a patient's risk profile and the quarter-year and hospital-year of birth. Figure A.6 Subfigure (a) shows a strikingly similar distribution of the physician fixed effects for male and female physicians, suggesting that neither male nor female physicians are more skilled. Next, we confirm that physicians are not being matched to team types based on skill. To do this, we first rank physicians by their skill (i.e., using their fixed effect from Equation A.5) and standardize by the total number of physicians to create a percentile distribution from 0 to 1, where lower ranks imply greater skill. In Figure A.6 Subfigure (b) we show that a two-member team of higher-ranked doctors (i.e., lower skill based on the solo sample) indeed have higher maternal complication rates. However, in Figure A.6 Subfigure (c) we show that the gender mix of the team does not appear to be correlated with physician skill. Specifically, we then show (i) the mean of the rank of the four team types (close to 0.5 for all team types), and (ii) the rank distance, i.e., the

difference in ranks between the Lead minus (-) the Assisting physician (about a 0-2 percentage point difference across team types). A negative rank distance means the Lead physician has a lower rank than the Assisting physician, i.e., the Lead is potentially more skilled than the Assisting.

F Mechanisms Analyses

F.1 How does gender mix affect team performance?

Here, we provide more details on the definitions and analyses from Section 4.1 of the manuscript.

F.1.1 Estimating Preferences

Individual physician preference for C-section: We estimate an individual physician's *preference* for C-section by estimating physician fixed effects using the solo sample (i.e., sample of deliveries attended by a single physician). For example, a physician's individual preference for a C-section is the physician fixed effect obtained from regressing C-section (indicator equal to 1 for C-section) on physician fixed effects, patient controls, quarter-year, and hospital-year fixed effects as in Equation 1 within the solo sample. This physician fixed effect captures how much a patient's likelihood of receiving a C-section changes if their baby is delivered by *Physician A* instead of *Physician B* in the solo sample while keeping constant patient, hospital, and time characteristics.

Team preference for C-section: We estimate a team's preference in a very similar way to individual preferences: by estimating team fixed effects for each practice using the team sample (i.e., sample of deliveries attended by a team) after regressing an indicator for the discretionary practice on team fixed effects, patient controls, quarter-year and hospital-year fixed effects as in Equation 1. However, to ensure we have a sufficient number of observations to calculate a team fixed effect, we limit the sample to teams that performed at least 35 births together. We use 35 as a cut-off because it is the same cut-off used at the individual-physician level (see our sample selection Figure A.1. This reduces the sample to 387,982 births (approximately 70% of the original team sample). This team fixed effect captures how much a patient's likelihood of receiving a C-section changes if their baby is delivered by *Team X* instead of *Team Y* while keeping constant patient, hospital, and time characteristics.

F.1.2 Association between C-sections and maternal complications

To establish the association between C-sections and higher rates of maternal complications, we run a "first stage" regression constructed as follows. We estimate the relationship between maternal complications and preferences for C-section by regressing a maternal complications indicator (equal to 1 if any of 25 severe maternal complications is recorded on the patient's record) on a standardized version of the physician or team fixed effect (as outlined above). Figure A.7 Panel A shows that as either an individual physician's or a physician team's preference for C-section increases, the likelihood of a maternal complication for the patient also rises. This supports our assumption that these practices are indeed correlated with worse patient outcomes and can plausibly be examined as a mechanism contributing to maternal complications.

F.1.3 The Gender Mix effect: An illustrative example using C-sections

The Gender Mix effect refers to the unique impact that the gender mix of the physician team has on a team's decision-making and patient outcomes. We use C-sections as a case study to provide evidence in support of the Gender Mix effect. As explained in Section 4.1, Figure V Panel A establishes that the physician's individual preferences for C-section map onto average team decisions and maternal complication rates (i.e., a two-member team of physicians who prefer C-sections perform more C-sections and have higher complication rates than a two-member physician team who prefer vaginal births). Figure V Panel B, then shows the same results by team type, and demonstrates that even within the same combination of member preferences, teams make different C-sections based on their gender mix. Below we provide a more detailed discussion.

The patterns in Figure V make intuitive sense. A patient is far more likely to receive a C-section when both team members prefer a C-section than when both team members prefer a vaginal delivery (i.e., compare the C-section rates between the "$Csection_L - Csection_A$" and the "$Vaginal_L - Vaginal_A$" preference combinations in Subfigures (a) and (c)). Moreover, when there is a conflict in preferences over delivery modes between the team members, the likelihood of a patient having a C-section becomes lower than a team that uniformly prefers a C-section but higher than a team that uniformly prefers a vaginal delivery. However, to study the Gender Mix effect, we are not interested in comparing estimates of C-section use and maternal complications *across* preference combinations; we are interested in comparing the four estimates *within* a preference combination. If the Gender Mix effect was NOT influencing C-section decisions across the four team types, then *within* each of the four combinations of preferences, the likelihood of choosing a C-section should be statistically indistinguishable across M_L–M_A, M_L–F_A, F_L–M_A, and F_L–F_A teams. However, we find the opposite: even when team members' preferences for delivery modes are held constant, the gender mix of the team still significantly affects the likelihood of performing a C-section.

Altogether, Figure V shows that differences in physician preferences for C-section does not fully explain the differences in performance across the four team types. Instead, the gender mix of the physician teams leads to significantly different use of C-sections even when the two physicians have the exact same preferences.

F.2 Why does gender mix affect team performance?

F.2.1 Gender mix affects the incorporation of individual preferences into team decisions

One potential and specific manifestation of how the gender mix of teams affects team performance is by changing how individual team members' preferences are aggregated and incorporated into team decisions. Appendix Figure VI shows that same-gender teams' joint preferences align closely with their members' individual preferences (the red and gray bars move in the same direction), while mixed-gender teams display notable divergence between individual and joint preferences (the red and gray bars move in opposite directions). Female-only teams, in particular, are the only team that do not default towards performing more C-sections than their members' individual preferences would suggest (i.e., they are the only team type with a negative grey bar).

F.2.2 Gender mix affects team resilience to challenges.

Gender norms may affect how resilient teams are to challenging conditions that could make collaboration more difficult. For example, same-gender teams may benefit from shared communication and

practice styles and fewer hierarchical tensions, leading to smoother collaboration. In contrast, mixed-gender teams may face more frictions because social norms around gender and authority in medicine can introduce tension, particularly when a female physician is in the lead role. Empirically, we study how performance differs by team gender mix when the team faces challenges that may make teamwork more difficult. For these analyses, we estimate the following triple interaction model:

$$\begin{aligned} y_{ijkht} = \boldsymbol{\beta}(\text{lead_fem}_j \times \text{samegender_assist}_k \times Var_i) \\ + \gamma X_i + \delta Z_j + \phi T_k + \alpha_t + \alpha_{hy} + \varepsilon_{ijkht} \end{aligned} \tag{A.6}$$

where y is the observed outcome for patient i under the care of Lead physician j and Assisting physician k in hospital h and quarter-year t, and $\times$ signifies a fully saturated interaction (i.e., $A \times B$ implies the estimation of the main coefficients on A and B as well as the interaction term between the two). Var_i signified one of four challenges to teamwork, describe below:

1) Var_i = Conflicting preferences for C-section (binary): An indicator equal to 1 when the physician's preferences for C-section conflict (one physician prefers C-section and one physician prefers vaginal birth). Therefore, no conflict means either both prefer vaginal birth or both prefer C-sections. See Equation 2 for details on how conflicting preferences are measured.

2) Var_i = Lead physician younger than assisting physician (binary): This variable is equal to 1 when the Assisting physician is 5 or more years older than the Lead physician (i.e., the Lead is notably younger than the assisting physician). This analysis excludes controls for physician age since the triple interaction term is based on physician age.

4) Var_i = Limited prior collaboration (binary): Since our data does not capture the full work histories of physician teams, we identify teams with limited familiarity using the following approach. "Limited prior collaboration" is an indicator equal to 1 when a physician team has cumulatively delivered at *most* two babies together within their first two recorded quarters of working together in the data (teams that delivered more than two babies in their first two quarters are set to missing). This threshold corresponds to the 5th percentile of team birth volume, though our results remain robust when using alternative cut-offs ranging from the 1st to the 25th percentile. Overall, this analysis compares team performance on deliveries when teams have low initial familiarity to deliveries when teams have more experience working together. This analysis excludes controls for a physician's cumulative previous collaborations since the triple interaction term is based on this variable.

3) Var_i = High levels of hospital strain (binary): To calculate hospital strain, we first calculate how many deliveries occurred within a hospital in a given quarter. Second, we create within-hospital quintiles of delivery volume (i.e., quintile 1 and quintile 5 identify the quarters with the lowest and highest level of strain for *that hospital* across our entire data period). Then, "high strain" is identified as quintile 5.

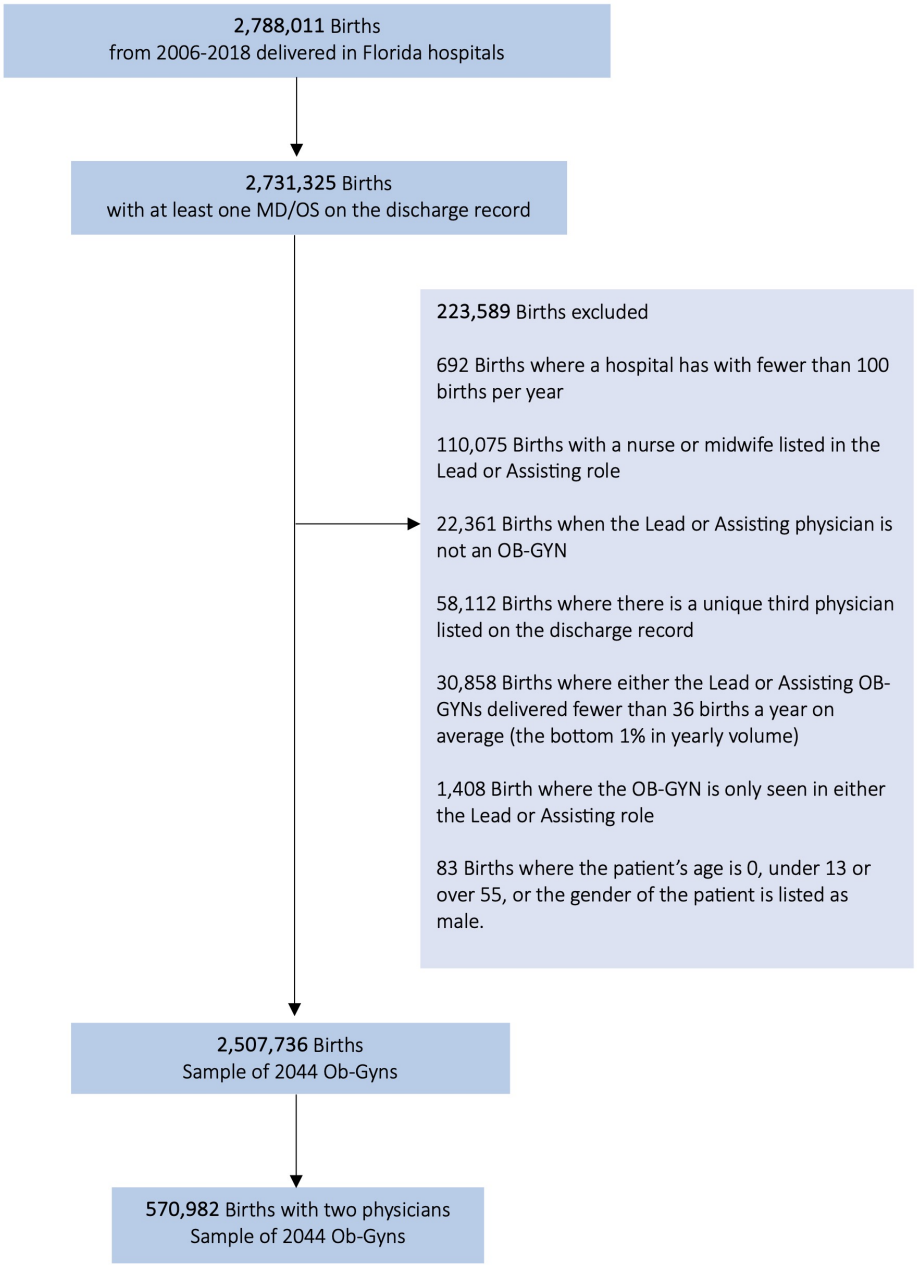
. Because hospital strain is calculated at the hospital-quarter level, we only include hospital and year fixed effects to have sufficient within-hospital variation.

Results of this estimation can be found in Figure VII. In addition to the results discussed in the main

manuscript, here we provide more discussion on why all teams except for F_L–M_A teams perform better when there is a conflict in preferences. When there is conflict in preferences between team members, then one physician in the team prefers a vaginal delivery-mode. This conflict lowers the likelihood of a M_L–M_A team, for example, of performing a C-section, and thus lowers the maternal complication rate for M_L–M_A teams with conflict. In short, "conflict" appears to exert downwards pressure on C-section use, leading to better outcomes except for F_L–M_A teams. As discussed in the manuscript, we speculate that this is because a female Lead with a male Assisting physician inverts traditional gender norms around leadership.

G Supplementary Figures and Tables

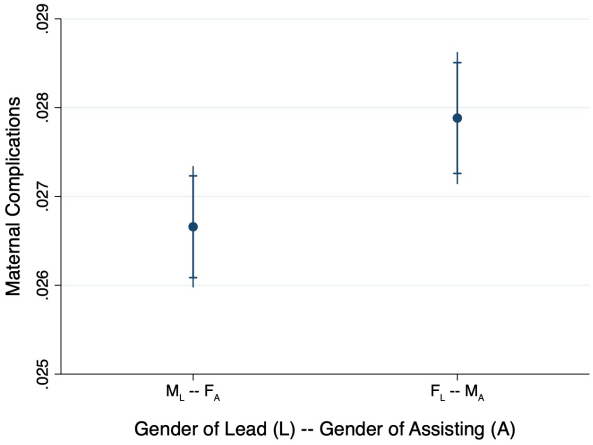
FIGURE A.1
Sample Criteria



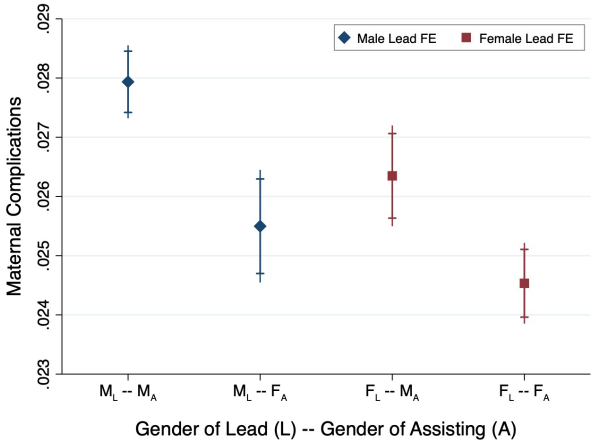
Notes: This figure presents the sample selection process to create the final analytic sample used in the analysis. Note that with the inclusion of some controls, such as a physician's cumulative births up to the prior quarter, the sample size is reduced to 540,400 births.

FIGURE A.2
Gender Mix and Maternal Complications: Within- Team or Within-Physician

(a) Within-team: Team FE



(b) Within-physician: Lead FE



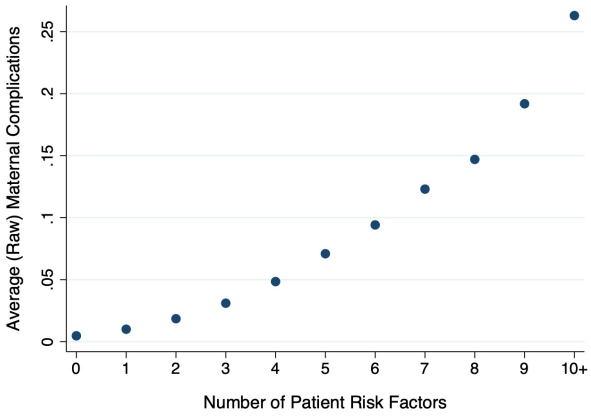
Notes: *M/F* indicates physician gender; subscript *L/A* denotes Lead or Assisting. *Maternal Complications* equals 1 if the patient experiences any of 25 adverse events during labor and delivery. Error bars depict 90% and 95% confidence intervals.

Subfigure (a) Fitted values from Appendix Equation A.1 with a fixed effect for each physician pair (role-invariant); therefore estimated only for mixed-gender teams.

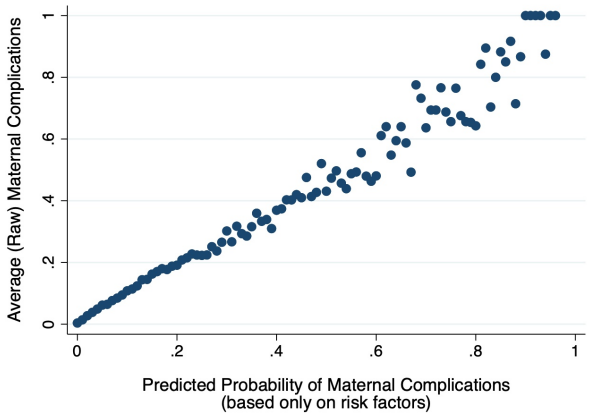
Subfigure (b) Fitted values from Appendix Equation A.2 with Lead-physician fixed effects, estimated separately for male Leads (blue diamonds) and female Leads (red squares).

FIGURE A.3
Risk Factors Predict Maternal Complications

(a) Number of risk factors



(b) Predicted maternal complication



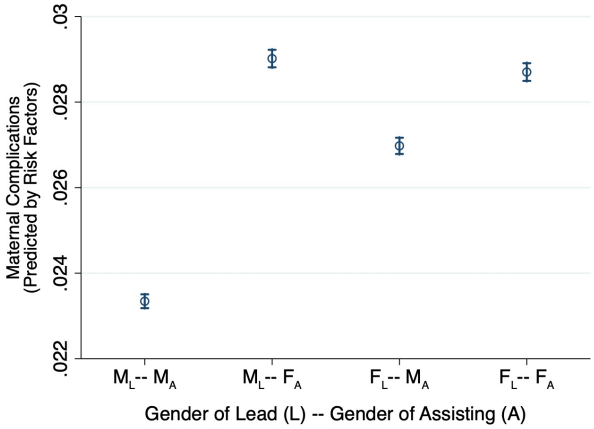
Notes: This figure shows how patient risk factors, observable to the physician before labor, relate to maternal complications.

Subfigure (a) Plots the unadjusted maternal complication rate against the raw count of patient risk factors.

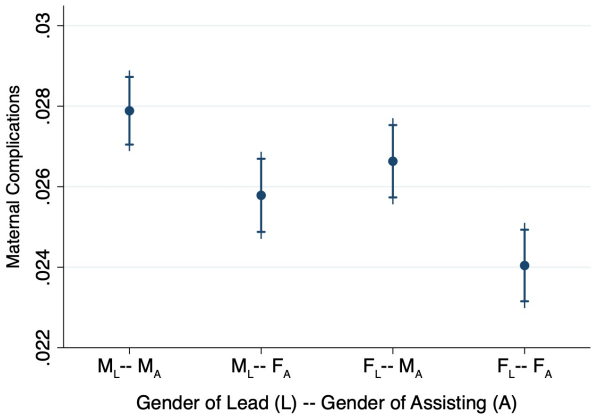
Subfigure (b) Plots the unadjusted maternal complication rate against the predicted probability of a complication, where the prediction comes from a regression of the complication indicator on 23 patient-risk indicators plus the number of diagnosis codes recorded in the encounter.

FIGURE A.4
Accounting for Coding Differences

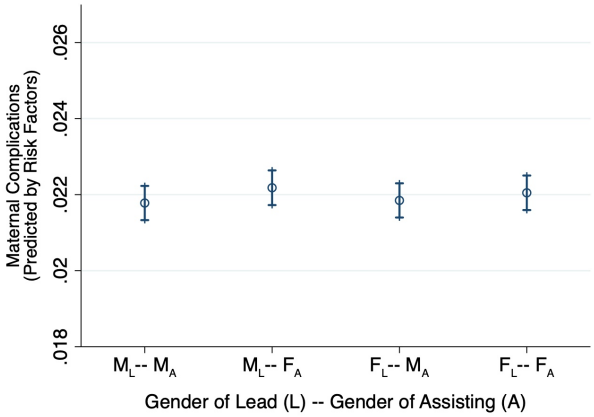
(a) Mean predicted rate of maternal complications based on risk factors ($\widehat{MC^{RF}}$), covariates as observed and num. diagnosis codes as observed



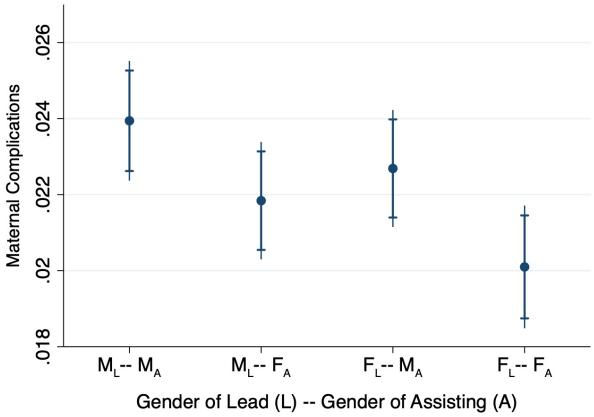
(b) Main regression estimates ($\widehat{MC}$), covariates set at their sample means and each indicator for num. diagnosis codes set at its sample mean



(c) Mean predicted rate of maternal complications based on risk factors ($\widehat{MC^{RF}}$), covariates as observed and num. diagnosis codes set at sample mean (=6)



(d) Main regression estimates ($\widehat{MC}$), covariates at sample mean and num. diagnosis codes set at the sample mean (=6)

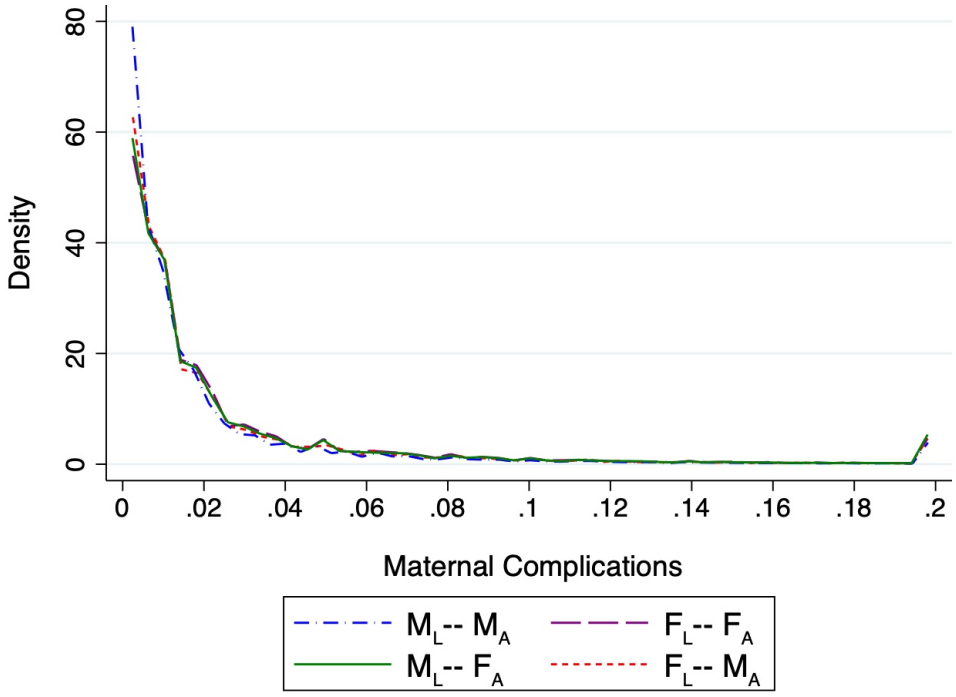


Notes: *M/F* denotes physician gender; subscript *L/A* identifies Lead or Assisting. *Maternal Complications* (MC) equals 1 if any of 25 adverse events occur during labor and delivery. Error bars show 90 % and 95 % confidence intervals.

Subfigures (a) & (c) Fitted values $\widehat{MC^{RF}}$ – mean predicted MC based solely on patient risk factors.

Subfigures (b) & (d) Fitted values $\widehat{MC}$ from Equation 1. In (b) each indicator for the number of diagnosis codes is set to its own sample mean; in (d) the total diagnosis-count variable is fixed at the overall mean (= 6) to remove coding-intensity differences across team types. See Appendix Section E for details.

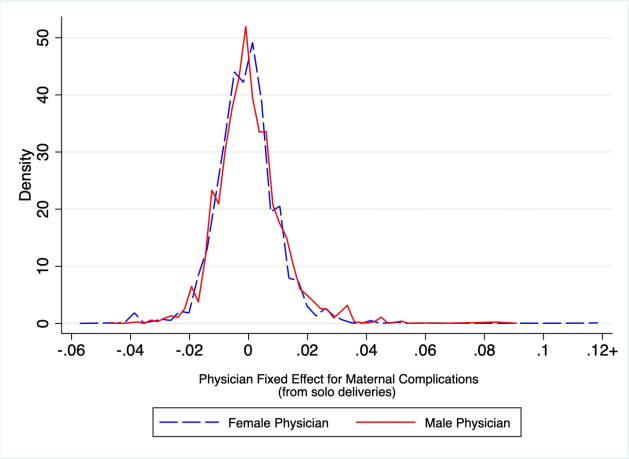
FIGURE A.5
Distribution of Predicted Maternal Complications by Team Type



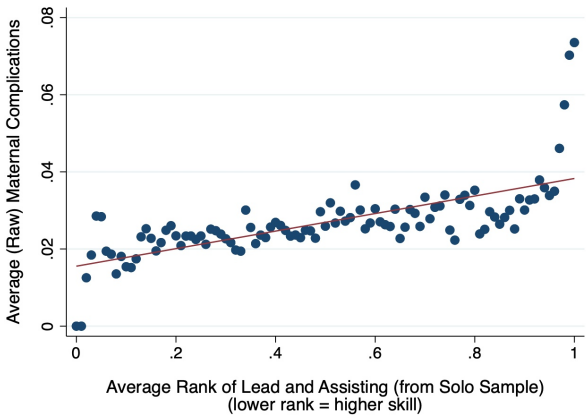
Notes: *M/F* indicates physician gender; subscript *L/A* denotes Lead or Assisting. *Maternal Complications* (MC) equals 1 if the patient experiences any of 25 adverse events during labor and delivery. Each line shows, by team type, the distribution of the predicted maternal complications $\widehat{MC^{RF}}$, obtained from a logistic regression of MC on 23 antepartum risk indicators and the number of diagnosis codes recorded for the encounter.

FIGURE A.6
No Differences in Physician Skill Across Team Types

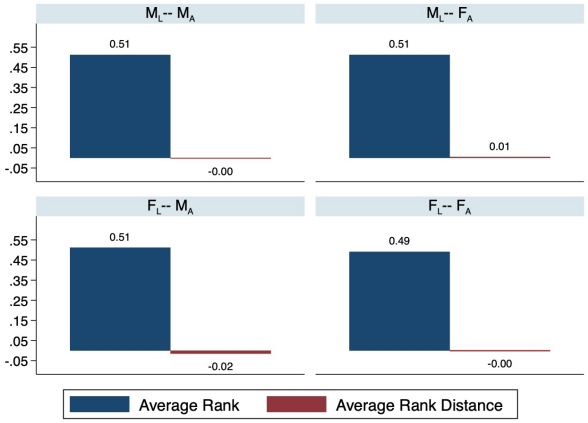
(a) Physician skill by gender



(b) Skill and maternal complications



(c) Skill within a team



Notes: *M/F* indicates physician gender; subscript *L/A* denotes Lead or Assisting. *Maternal Complications* (MC) equals 1 if the patient experiences any of 25 adverse events during labor and delivery.

Subfigure (a) Physician fixed effects from Appendix Equation A.5, estimated on "solo" births.

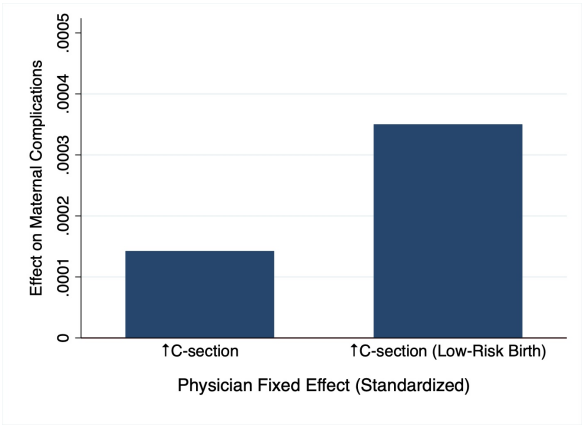
Subfigure (b) Relationship between the team's average physician skill rank (0 = highest skill, 1 = lowest skill) and the rate of maternal complications.

Subfigure (c) Average skill rank and rank distance between Lead and Assisting physicians across the four team types. See Appendix Section E for details.

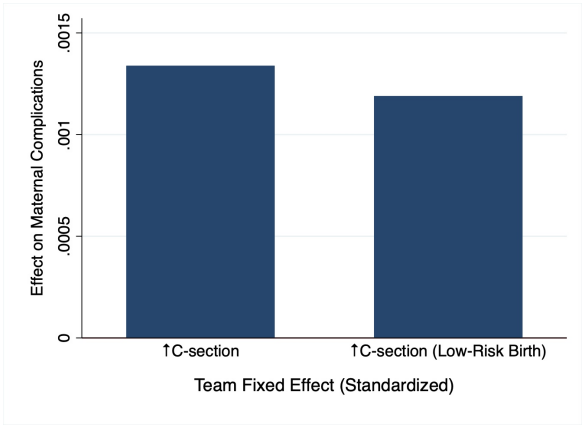
FIGURE A.7
C-sections and Maternal Complications

Panel A. C-section preferences and maternal complications

(a) Solo sample

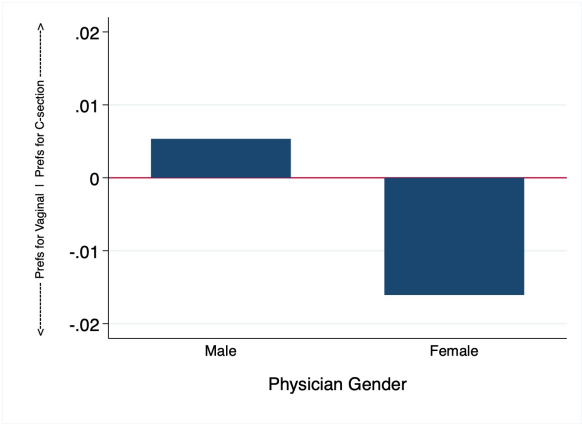


(b) Team sample

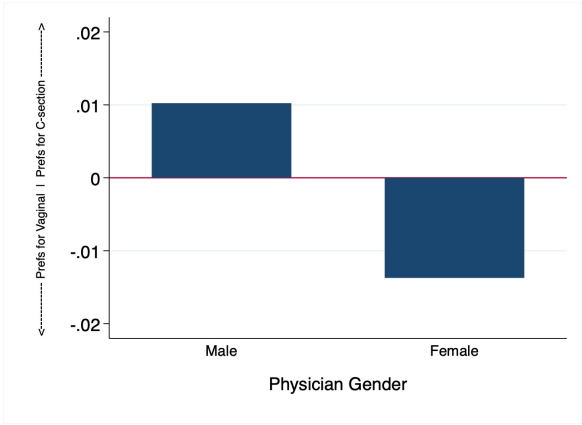


Panel B. Individual C-section preferences by physician gender

(c) All births

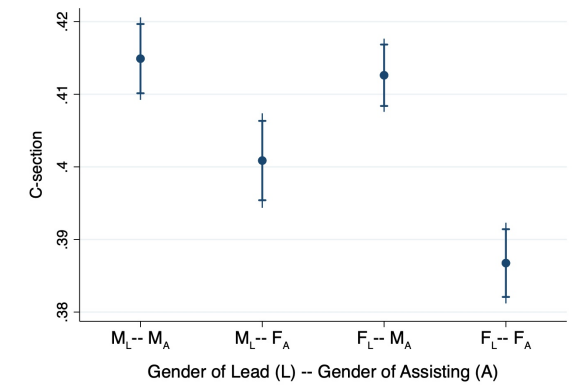


(d) Low-risk births

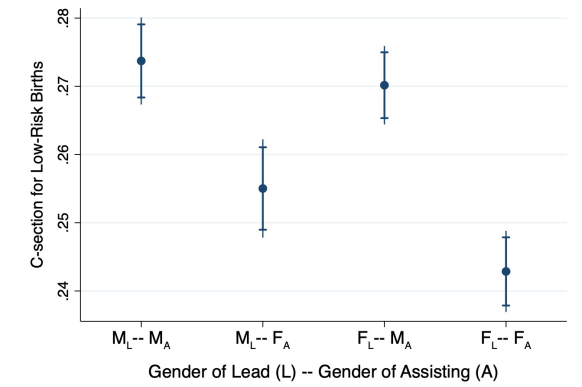


Panel C. C-section use by team types

(e) All births



(f) Low-risk births



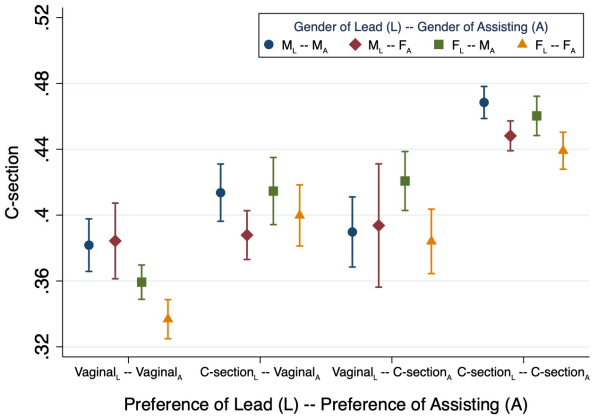
Notes: *M/F* indicates physician gender; subscript *L/A* denotes Lead or Assisting. *Maternal Complications* equals 1 if any of 25 severe events are recorded during labor and delivery. Error bars show 90% and 95% confidence intervals.

Panel A presents the correlation between individual and team preferences for C-section and maternal complications. Maternal complications are regressed on standardized (a) **physician** C-section preference (solo fixed effect) and (b) **team** C-section preference (team fixed effect).
Panel B presents mean physician C-section preference by gender for (c) all births and (d) low-risk births.
Panel C presents fitted likelihood of C-sections by team type using Eq. 1 for (e) all births and (f) low-risk births.

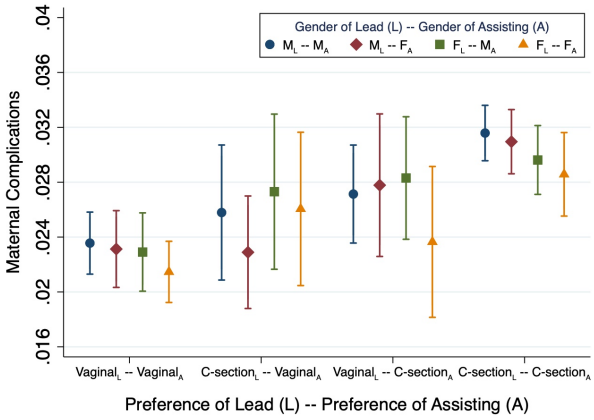
FIGURE A.8
The "How:" Exploring Mechanisms via C-section Decisions (Robustness Checks)

Panel A. Robustness to using terciles for preferences

(a) Team decision: C-section

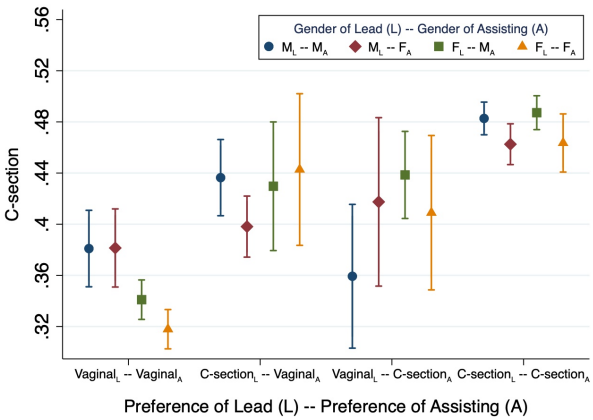


(b) Team performance: Maternal complications

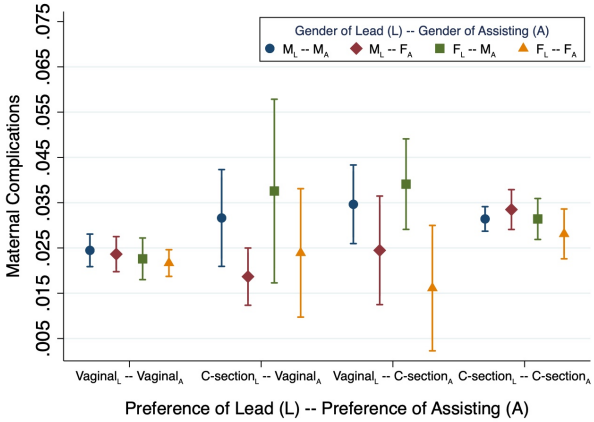


Panel B. Robustness to using quintiles for preferences

(c) Team decision: C-section



(d) Team performance: Maternal complications



Notes: *M/F* indicates physician gender; *Vaginal/C-section* denotes each physician's preferences based on physician C-section FEs (top quintile or tercile = C-section, bottom quintile or tercile = vaginal). Subscript *L/A* marks Lead or Assisting. Note the different Y-axis scales between the panels. Error bars show 90% and 95% confidence intervals.

Panel A shows fitted values from Equation 2 for varying preference combinations, using tercile cutoffs for preferences.
Panel A shows fitted values from Equation 2 for varying preference combinations, using quintile cutoffs for preferences.
Subfigures (a) and (c) use *C-section* as the outcome (1= C-section, 0= vaginal).
Subfigures (b) and (d) use *Maternal Complications* as the outcome (1= any of 25 severe events during delivery).

TABLE A.1
Summary Statistics: Physicians

	M_L–M_A	M_L–F_A	F_L–M_A	F_L–F_A	Full Sample
Lead Physician Characteristics					
Physician age	51.22	51.01	41.53	41.24	46.77
Births per year	201.41	200.16	159.00	156.63	181.59
Number of unique pairs (year)	7.27	6.48	6.81	7.85	7.11
Number of unique pairs (qtr-year)	2.28	2.19	2.22	2.42	2.28
Assisting Physician Characteristics					
Physician age	50.92	41.30	50.52	41.06	46.41
Births per year	200.23	157.05	199.63	156.35	180.35
Observations	201,164	128,695	107,955	133,168	570,982

Notes: *M/F* indicates physician gender; subscript *L/A* denotes Lead or Assisting. Table summarises physician characteristics in the team sample (births with two physicians). Unique pairs count the contacts per year and per quarter for each distinct Lead–Assisting pair. **Sample size**: 2,044 physicians (1,010 female Leads; 1,034 male Leads) delivering 570,982 team births. Female physicians serve as Lead in 42% of births; male physicians in 58%.

TABLE A.2
Robustness: Logistic Regression

	(1) Main Specification	(2) Logistic Regression
Female Lead	0.0008	1.0316
	(0.0008)	(0.0468)
Same Gender Assisting	0.0021***	1.1610***
	(0.0008)	(0.0396)
Female Lead × Same Gender Assisting	-0.0047***	0.7602***
	(0.0011)	(0.0372)
Physician Controls	Yes	Yes
Patient Controls	Yes	Yes
Hospital Fixed Effects	No	Yes
Quarter-Year Fixed Effects	Yes	Yes
Hospital-Year Fixed Effects	Yes	No
Ymean	0.026	0.026
Observations	540400	540411
R^2	0.093	0.179

Notes: *Maternal Complications* equals 1 if a patient experiences any of 25 adverse events during labor and delivery. *Female Lead* equals 1 if the Lead physician is female; 0 if male. *Same-Gender Assisting* equals 1 if the Assisting physician matches the Lead's gender; 0 otherwise. Significance levels: $p < .01^{***}, p < .05^{**}, p < 0.1^{*}$.

Column 1 reports regression coefficients with standard errors (in parentheses) from Equation 1.

Column 2 reports odds ratios and the pseudo-R^2 from a logistic regression that includes quarter-year and hospital fixed effects (more detailed interacted fixed effects prevent convergence).

TABLE A.3
Robustness: Varying Samples

	(1) Drop Planned or Prior C-Section	(2) Drop Maternal Fetal Medicine Specialists	(3) Drop Physicians Under Age 33	(4) Drop Teaching Hospitals	(5) Drop Hospitals Without All Team Types
Female Lead	0.0009	0.0009	0.0012	0.0015*	0.0004
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
Same Gender Assisting	0.0018**	0.0026***	0.0022***	0.0026***	0.0019**
	(0.0008)	(0.0007)	(0.0008)	(0.0007)	(0.0008)
Female Lead × Same Gender Assisting	-0.0039***	-0.0051***	-0.0050***	-0.0061***	-0.0042***
	(0.0011)	(0.0011)	(0.0012)	(0.0012)	(0.0012)
Ymean	0.023	0.025	0.026	0.029	0.025
Observations	414311	506844	478516	381613	475492
R^2	0.077	0.088	0.094	0.082	0.094

Notes: Regression coefficients with standard errors (in parentheses) from Equation 1. *Maternal Complications* = 1 if any of 25 adverse events occur during labor and delivery. *Female Lead* = 1 if the Lead physician is female; 0 if male. *Same-Gender Assisting* = 1 if the Assisting physician matches the Lead's gender; 0 otherwise. Significance levels: $p < .01^{***}, p < .05^{**}, p < 0.1^{*}$.

TABLE A.4
Robustness: Different Outcome Definitions

	(1) Main MC Measure (2006-2015Q3)	(2) Expanded MC Measure (2006-2018)	(3) Exclude Blood Transfusions (2006-2018)
Female Lead	0.0010	0.0017*	0.0018**
	(0.0008)	(0.0009)	(0.0007)
Same Gender Assisting	0.0016**	0.0021**	0.0008
	(0.0007)	(0.0008)	(0.0006)
Female Lead × Same Gender Assisting	-0.0039***	-0.0054***	-0.0028***
	(0.0012)	(0.0013)	(0.0010)
Ymean	0.019	0.035	0.021
Observations	380675	540400	540400
R^2	0.100	0.124	0.084

Notes: Regression coefficients with standard errors (in parentheses) from Equation 1. *Female Lead* = 1 if the Lead physician is female; 0 if male. *Same-Gender Assisting* = 1 if the Assisting physician matches the Lead's gender; 0 otherwise. Significance levels: $p < .01^{***}, p < .05^{**}, p < 0.1^{*}$.

Column 1 *Maternal Complications* (MC) = 1 if any of the CDC's 25 severe-maternal-morbidity indicators is recorded, using only quarter-years before the ICD-10 transition.

Column 2 MC = 1 if any of the CDC indicators *or* the additional indicators in Bateman et al. (2013) is present.

Column 3 Same definition as Column 2 but excluding cases identified solely by blood transfusion codes.

See Appendix Section C for variable details.

TABLE A.5
Risk Factors and Maternal Complications

	Maternal Complications	
	Odds Ratio	S.E.
Patient Age		
Age (≤22)	1.152***	(0.018)
Age (23–26)	1.009	(0.016)
Age (27–30)	1.001	(0.015)
Age (31–33)	1.027	(0.016)
Asthma	0.529***	(0.012)
Poly- & Oligo-hydramnios	0.746***	(0.016)
Maternal physical abnormality	0.733***	(0.011)
Blood disorders	2.474***	(0.037)
Uterine size issue	0.799***	(0.047)
Infant size issue	0.796***	(0.015)
Obesity	0.439***	(0.008)
Anemia	2.967***	(0.033)
Diabetes	0.691***	(0.012)
Substance abuse or smoking	0.448***	(0.007)
Infectious and parasitic conditions	0.517***	(0.011)
Heart disease	1.561***	(0.043)
Known fetal abnormality	0.737***	(0.023)
Hypertension	1.345***	(0.016)
Isoimmunization	0.746***	(0.024)
Other conditions/risks	1.291***	(0.023)
Ruptured membrane	0.696***	(0.015)
Previous pregnancy	0.982	(0.013)
Previous C-section	1.424***	(0.016)
Breech	1.074***	(0.016)
Multiple gestation	1.387***	(0.032)
Pre-term birth	1.283***	(0.019)
Observations	2507736	
R^2	0.193	

Notes: *Maternal Complications* equals 1 if a patient experiences any of 25 adverse events during labor and delivery. Each cell reports the coefficient β^r from Equation A.3, giving the patient's predicted probability of a complication as a function of the listed risk factor. Age coefficients are relative to the reference group, patients aged 34 and older. Indicators for the number of diagnosis codes are included in the regression but omitted from the table for brevity.Significance levels: $p < .01^{***}, p < .05^{**}, p < 0.1^{*}$.