The Role of Physician Training in Racial Disparities in Maternal Healthcare

Helen Kissel* and Helena Roy†‡

November 2, 2025

Abstract

Black mothers in the United States face disproportionately high rates of pregnancy complications, such as pre-eclampsia, and face a maternal mortality rate more than twice that of non-Hispanic White mothers. A potential instrument for reducing disparities is training physicians in how to care for patients of different races. Residency provides the bedrock for specialist practice in obstetrics and gynecology, and could be a valuable training period in which to learn race-specific risks and considerations. In this paper, we examine whether racial disparities in treatment use and outcomes can be traced back to physicians' residency training, as reflected in the practice patterns of their alumni. We find that residency programs have only modest effects on variation in treatment use and outcomes across racial groups. Notably, a residency's relative effect on C-section use and delivery complications is not strongly correlated across Black versus White patients, suggesting non-uniform learning across different patient groups. We assess if training at a residency with a more diverse patient mix, a feature advertised by multiple residencies as an important part of physician education, impacts physicians' deviation from clinical consensus in delivery decisions. Though we document that deviations from clinical consensus are more likely for Black patients, we find essentially no impact of training with a diverse patient population on this racial gap.

JEL Classification: I14, I23

*Kissel: Stanford University – hkissel@stanford.edu

†Roy: Hasso Plattner Institute – helena.roy@hpi.de

‡We are grateful to our advisors Ran Abramitzky, Liran Einav, Petra Persson, Maya Rossin-Slater, and Heidi Williams for their extensive support and advice. We thank Harsh Gupta, Adrienne Sabety, Jason Weitze, as well as participants of seminars at Stanford University for valuable feedback. We are grateful for the expertise and guidance of several obstetricians, gynecologists, midwives and labor and delivery nurses, especially Elliott Main. This research was supported by a Stanford University's Department of Population Health Sciences via a Spectrum Research Grant.

Data for this project were accessed using the Stanford Center for Population Health Sciences Data Core. The PHS Data Core is supported by a National Institutes of Health National Center for Advancing Translational Science Clinical and Translational Science Award (UL1TR003142) and from Internal Stanford funding. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

1. Introduction

Maternal health in the United States (US) is poor relative to peer countries, with maternal mortality rates two to three times higher than other high-income nations (Gunja et al., 2024). Outcomes differ substantially by race: the maternal mortality rate for Black women is 2.5 times higher than for White women, and racial gaps in outcomes persist across income or education gradients (Corredor-Waldron et al., 2024; Gunja et al., 2023; Hoyert, 2022; Kennedy-Moulton et al., 2022). Black patients experience a range of complications at different rates from White patients, suggesting that accounting for race in diagnosis and treatment is important.[1] But there is evidence that the use of treatments and interventions such as C-sections is less tailored to Black patients' clinical indications (Corredor-Waldron et al., 2024; Robinson et al., 2023). This has led to substantial policy interest in how to reduce racial disparities and improve outcomes.

A key potential policy instrument is physician education. Providing physicians with structured training on racial disparities could strengthen their ability to make equitable clinical decisions and respond appropriately to race-specific risks. Similar effects could plausibly be achieved by ensuring physicians train with a diverse patient population, to ensure exposure to a range of patient profiles and learn important heterogeneity through experience. Residency provides the bedrock for physician practice in specialties such as obstetrics and gynecology (hereafter, OB/GYN). The sizable geographic variation in practice style (such as C-section use), patient population, and outcomes could have long-run implications if such variation influences the practice styles of alumni.

In this paper, we estimate the influence of residency on physician performance, focused on racial disparities in C-section use and complication rates. We ask two questions. First, do residencies produce physicians with the same relative treatment use and performance across patients of different races? Second, can the racial patient mix at a residency—in other words, training with a more diverse group of patients—improve a physician's ability

[1] Indeed, this is captured in national clinical guidance, such as the US Preventive Services Task Force's recommendations for low-dose aspirin use during pregnancy, which highlights higher incidence of pre-eclampsia among non-Hispanic Black women and includes Black race as a risk factor (US Preventive Services Task Force et al., 2021).

to treat patients of different races more equitably in post-residency practice? We measure residencies' relative contribution to each of these outcomes via a fixed effect, controlling for a range of patient characteristics and exploiting quasi-random variation in patient-physician assignment at delivery. By measuring the relative performance of physicians from different residencies within a hospital, we mimic an experiment where patients are randomly allocated to a residency alum upon arrival in a labor and delivery ward. We then examine the overlap between the residencies' relative ranking and performance across Black and White patients.

There are two primary reasons it might be important for physicians to consider patient heterogeneity by race. The first is clinical: many medical conditions differ in their incidence or manifestation depending on race,[2] and medical education has historically focused on the typical presentation among White patients.[3] If physicians form diagnostic priors based on typical incidence or manifestation among White patients, they may systematically misdiagnose other groups at higher rates.[4] The second is relational: a legacy of discrimination and marginalization has contributed to lower levels of trust in medical providers among some racial and ethnic groups (Alsan and Wanamaker, 2018). Physicians who recognize and respond to these dynamics by acknowledging concerns and building trust, may achieve better outcomes through more effective communication and positive patient relationships.

A potential contributor to racial disparities may therefore be physicians' inability to account for heterogeneity when treating patients of different races. How might this channel be mitigated? Some have focused on diversifying the physician workforce, with the idea that race-concordant physicians may be better able to treat patients from racial minorities. Indeed, concordant physicians are better able to generate patient engagement with preventative healthcare (Alsan et al., 2019) and can improve outcomes for patients of the same race

[2]For instance, Black pregnant patients have higher incidence of particular maternal complications—such as preeclampsia and other hypertensive disorders—as well as worse outcomes (Fingar et al., 2017).

[3]For example, a study of medical images printed in the New England Journal of Medicine from 1992-2017 found that over 80% were of White patients, an outsize share relative to the US population (Massie et al., 2021). Moreover, much of the clinical evidence underlying treatments is performed on predominantly White populations (Alsan et al., 2024).

[4]In the setting of stroke diagnosis, for example, (Philip and Ozkaya, 2024) finds a large Black-White gap in diagnosis rates stemming largely from the decision to test. Disparate treatment can be disaggregated into differences in thresholds for diagnosis across race, and the *unjustified skill gap*. On the latter, Philip and Ozkaya, 2024 find that physician error systematically disfavors Black patients. They write that this "can be interpreted to be a consequence of low physician effort, the use of incorrect priors and stereotypes, or of race-insensitive medical protocols."

or gender (Greenwood et al., 2018). Ensuring the clinical workforce is representative of the population is a worthy goal, but a long-term solution at best. Physicians in the US remain overwhelmingly White (64%) and male (63%), and the pipeline to qualification is long (AAMC, 2023). Moreover, concordance is no guarantee of improved outcomes. Indeed, recent findings from Corredor-Waldron et al., 2024, found that racial gaps in unscheduled C-sections persist even among Black physicians, though the gap appears smaller for Black physicians than for White physicians. A strategy adopted by some residencies is to train physicians to recognize and account for heterogeneity by demographic characteristics through coursework on health equity and clinical experience with diverse patient populations. We show two examples of this for OB/GYN residencies in Figure 1.

Childbirth offers a useful setting to examine these dynamics for several reasons. First, childbirth is one of the most common reasons for inpatient hospitalization in the US, generating large sample sizes. Second, delivery is typically overseen by a single physician who exercises substantial influence over treatment choice, meaning decisions and outcomes can be attributed to individual providers. Finally, the rotational variation in physician shifts and unpredictable onset of labor generate quasi-random assignment of patients to physicians, limiting concerns of patient selection into physicians based on unobservable preferences or risk.

We use Medicaid claims data from 2015-2019, covering 42% of all births. We supplement this with data on physician demographics and education from the American Medical Association's (AMA) MasterFile, American Community Survey (ACS), and the Centers for Medicare and Medicaid Services' (CMS) National Plan and Provider Enumeration System (NPPES). For all births, we focus on patients without scheduled C-sections, such that the delivery decision is made after a trial of labor by the physician attending birth.[5] Due to rotational variation in physicians' hospital shifts and the unpredictable onset of labor, low-risk patients in our sample are effectively assigned to whichever OB/GYN happens to be on shift the day they go into labor. Our empirical strategy mimics a natural experiment, where patients are randomly assigned to a residency alumni upon arrival at the hospital. This allows us to estimate a fixed effect for each residency, controlling for the patient's observable health risk

[5]For the outcome of C-section, we specifically focus on nulliparous, singleton, term, and vertex (NTSV) births.

and the hospital and month-year of delivery. This gives us a measure of how likely physicians from a given residency are to perform a C-section, or experience a complication, relative to other physicians treating clinically similar patients in the same hospital.

We find that across treatment use and patient outcomes, for both Black and White patients, the impact of a physician's residency training is small. This suggests that—in spite of substantial geographic variation in practice style, patient mix, and other characteristics shaping the residency environment—differences in hospital practice environment during residency do not translate to long-term differences in physician practice style and performance. Further, for C-section use and delivery complications, there is a positive but weak relationship between a residency's relative effect on White versus Black patients. The weak correlation suggests that the skills or practices acquired during residency may not generalize uniformly across patient populations.

Given the sizable racial disparities in maternal health and the lack of overlap between estimated residency effects for White and Black patients, our second research question asks whether physicians who train in areas with a higher share of Black patients can achieve more equal outcomes across patients of different races. We use geographic variation in population composition of a residency's local area to measure exposure to different patient groups during clinical education. We examine the impact of this variable on the racial gap in deviations from the clinical consensus in the diagnosis of delivery method. Our results suggest diagnostic deviations are more likely for Black patients (even within hospital), but training with a more diverse patient mix has little discernible impact on the racial gap. These findings indicate that while diversity in clinical training environments may be valuable, there is no measurable impact on how physicians treat patients of different races.

Our work relates to several literatures within economics and medicine. Most closely related is the literature in health economics on maternal healthcare, which has examined the determinants of physician decision-making and maternal health outcomes (Corredor-Waldron et al., 2024; Currie and MacLeod, 2017; Johnson and Rehavi, 2016; Kennedy-Moulton et al., 2022; Robinson et al., 2023). In both the economic and medical literature, other characteristics of medical school and residency—such as ranking (Asch, 2009; Schnell and Currie, 2018) and intensity of healthcare utilization (Phillips et al., 2017)—have been found to have long-lasting consequences on the style and quality of treatment choices, as well as patient outcomes. Prior studies examining whether residency hospital practice styles predict physicians' later C-section rates find positive but weak associations, suggesting limited persistence of training effects (Dranove et al., 2011; Epstein and Nicholson, 2009).

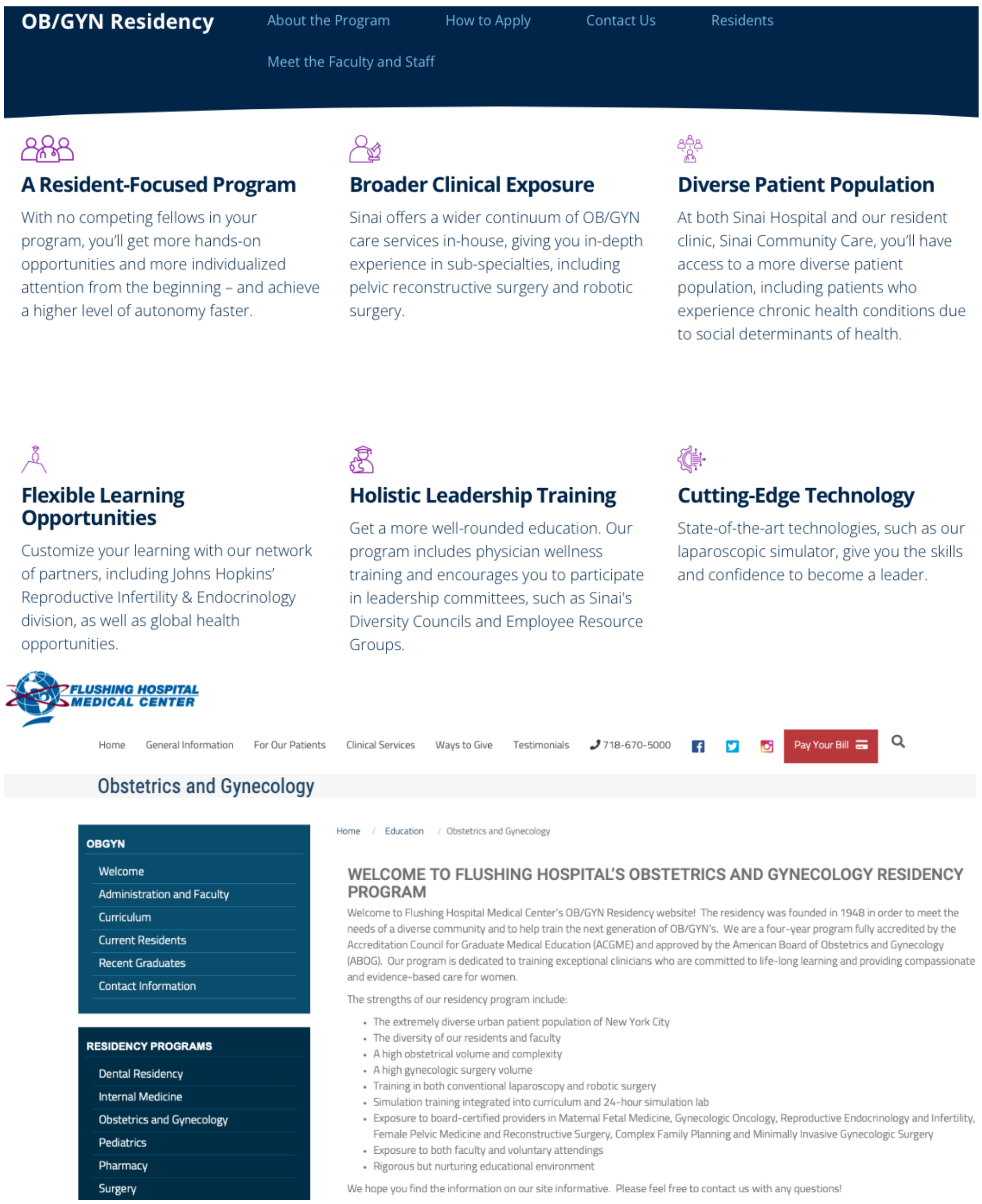
Our work also relates to research across innovation and health, showing that exposure to different demographic groups can shape attention to those groups' needs and outcomes. In innovation, Einiö et al., 2022 find that creators develop products more likely to be purchased by demographically similar users and that exposure to users with specific characteristics (such as low income) increases the likelihood of designing for their needs. In health, Gørtz et al., 2024 show that female patients of male physicians with daughters are less likely to die from female-specific cancers; an effect linked to greater adherence to female health guidelines, collaboration with female colleagues, and improved patient trust and communication. The rest of the paper proceeds as follows. In Section 2 we provide additional background on maternal health and training in OB/GYN, and present our data sources alongside sample descriptives. We outline our benchmark empirical strategy in Section 3, and provide results inSection 4. We conclude in Section 5 with a discussion of our results and other potential mechanisms driving physician-level variation in patient outcomes by race.

2. Background and Data

2.1 Background in Obstetrics and Gynecology

Our setting is obstetrics and gynecology (OB/GYN). Care in obstetrics covers pregnant patients during pregnancy, through delivery, and the postpartum period. Typically, a woman who discovers she is pregnant will register with an OB/GYN clinic and receive prenatal care throughout her pregnancy. This will include regular check-ups, with diagnostic tests such as scans and blood tests to monitor the mother's and infant's health, as well as planning for birth. Patients will then deliver, an inpatient procedure usually performed at a hospital, during which they will either receive a C-section or have a vaginal birth. The mode of delivery will be decided through discussions with physicians both before or during labor. For

Figure 1: Examples of OB/GYN Residency Statements on Patient Population

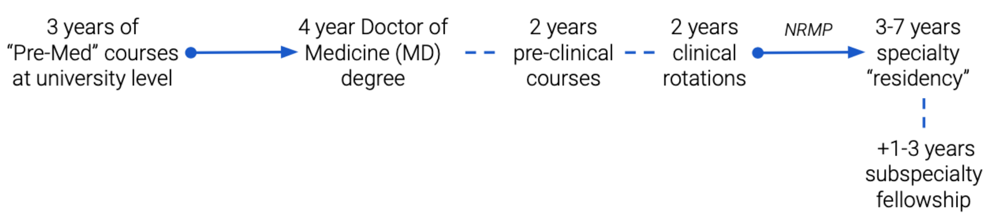


Notes: Examples from two OB/GYN residencies explicitly advertising a diverse patient population on their websites. Images taken as of August 2024.

example, some women may receive a scheduled C-section based on clinical factors determined in prenatal care (such as placenta previa or a previous C-section). The majority of women will enter labor spontaneously or via induction and attempt a vaginal delivery. Some patients will receive an emergency/unplanned C-section after a trial of labor, based on the physician's assessment of the risks and the progression of labor.

OB/GYN is a medical specialty that requires specific residency training. This training happens in a residency after medical school. Residencies are four years in length. In Figure 2 we outline the general medical training pathway in the US. After residency, physicians will typically join a clinical practice, at which they will both deliver prenatal care in-clinic (outpatient work) and have shifts at a hospital performing deliveries (inpatient work). It is worth noting that while patients may see one physician more than others during their prenatal care, matching with a physician for all prenatal care can be challenging due to the mixed nature of a physician's work, both in the clinic and in the hospital. If an issue arises during pregnancy, a given physician may not be available (e.g., if they are performing deliveries at that time). In interviews with OB/GYNs, we heard that many practices are moving towards a model where patients intentionally rotate physicians during prenatal care appointments, so as to see and meet a range of providers.

Figure 2: General Path for Physician Medical Training



Notes: Diagram shows the typical path for medical training of physicians in the US, covering all medical specialties. The National Residency Match Program is abbreviated to NRMP.

2.2 Data

Our primary data source for recording physician choices and patient outcomes is Medicaid claims data from 2015 through 2019. Medicaid covers approximately 42% of births in the US. We link these claims data to information on physician demographics and training from the AMA MasterFile for 2014 and the Center for Medicare and Medicaid Services' (CMS)

National Plan and Provider Enumeration System (NPPES). Finally, we include information from the American Community Survey (ACS) on the demographic composition of residency ZIP codes to measure residency patient mix.

The Medicaid claims data allow us to measure patient characteristics, procedures, and outcomes during pregnancy, delivery, and postpartum. We identify births based on ICD-9 and ICD-10 diagnosis codes following Auty et al., 2024. One key outcome is delivery method (C-section or vaginal birth). To control for clinical comorbidities that influence delivery method, we use a combination of ICD-9 and ICD-10 diagnosis codes from Gregory et al., 2002, Asch, 2009, Currie and MacLeod, 2017, and Robinson et al., 2023. These identify a patient's clinical indications for C-section appropriateness and risk factors, including asthma, oligohydramnios and polyhydramnios, preeclampsia and eclampsia, macrosomia, parity, singleton, term, and vertex.

Our second key outcome of interest is delivery complications. We combine two metrics of delivery complications. The first was developed by Asch, 2009, based on clinical advice, to study the association between a physician's residency ranking and their patient outcomes in Florida. The second was developed by HealthGrades, Inc., to assess the quality of maternity care across hospitals in the US (Turner et al., 2010). For vaginal deliveries, these include laceration, hemorrhage, thrombotic complications, among others. For C-sections, these include hemorrhage, infection, or operative and thrombotic complications. We make adjustments to this measure based on the frequency of certain complications as well as the degree to which a physician is responsible. For example, we exclude anesthetic complications from the measure, as the delivering physician is not responsible for administering anesthesia. We detail the process of constructing this measure, as well as all of the diagnosis codes used, in Appendix A.

For information on physician education and residency, we use the AMA MasterFile for 2014 and the CMS NPPES. Collectively, these provide information on physician demographics (such as gender and age) as well as residency location and graduation year. We restrict our analysis to physicians who completed their residency between 1989 and 2014. This means all included physicians should have between one and thirty years of practice experience post-residency when we observe outcomes in the Medicaid claims data.

To measure the racial patient mix at residency, we compute the share of the female population of childbearing age that is Black from the ACS 2011. We use the residency's 3-digit ZIP as the geographic area from which patients are drawn. We follow Alsan et al., 2024, who compute a similar measure of the share of a physician's patients who are Black, using estimates from the ACS at the 5-digit ZIP level.

2.3 Sample Restrictions

After identifying births from the Medicaid claims files, we have a sample of 7,831,460 deliveries. We make several further restrictions to this. First, we merge in the AMA MasterFile for 2014. This successfully matches with 29,700 distinct providers on 4,380,218 deliveries (a match rate of 56% for deliveries and 24% for physicians). This rate is driven in part by some claims not having an individual physician associated with the delivery (for example, a hospital NPI might be used to fill multiple provider fields); as such, there is no physician to match with in our primary file on outcomes. Additionally, because we are using data from 2014, we are only able to link physicians who had completed residency by 2014. Moreover, comparison with other national databases, such as the NPPES, has shown that while the AMA MasterFile is extremely comprehensive, it does not contain a complete census of physicians (Badinski et al., 2024). For comparison, Badinski et al., 2024 successfully match 43% of providers listed on Medicare claims—which have more comprehensive provider information—to the AMA MasterFile for 2014. And while our physician match rate is relatively low, the physicians identified are responsible for a disproportionate share of deliveries, suggesting we are identifying high-volume practitioners.

We further restrict to physicians who are responsible for at least 100 deliveries in Medicaid over the five years in our sample, and who attended residency between 1989 and 2014 (such that they have between one and twenty-seven years of post-residency practice experience at the start of our sample). This leaves us with 3,098,437 deliveries to 10,397 distinct providers. Then, we restrict to claims for deliveries with an identified hospital (as opposed to those associated with a physician group practice, or a birth center). Finally, we drop scheduled C-sections from our sample, as these interfere with the quasi-random variation in physician-patient assignment that we use in our empirical strategy.

In analyses of C-section usage, we restrict further to nulliparous patients only; in other words, patients who have had no children prior to the delivery. This is because the delivery method for higher-order births is typically determined by the method of prior deliveries. This makes it challenging to interpret decisions as those of the delivering physician in higher-order births (as there is also path dependence from the delivering physician's choice in the prior delivery).

2.4 Measuring C-Section Risk

In our analyses, we use a patient's C-section risk both as a control and to generate measures of a physician's deviation from clinical consensus in delivery method. We compute a patient's C-section risk using a logistic model. For every patient i:

$$\mathbb{P}(\text{C-section}_i = 1) = F(\beta X_i) \tag{1}$$

We include a large number of clinical indications in X_i, drawing on Gregory et al., 2002, Asch, 2009, Currie and MacLeod, 2017, and Robinson et al., 2023 for a comprehensive list of conditions and ICD codes.[6]. These covariates are listed in Table A4, with comparisons to previous literature.[7]

We estimate C-section risk on our entire sample (comprising 7,831,460 deliveries over 2015-2019). Collectively, the deliveries are performed by over 20,000 physicians: as such, the estimates are insensitive to any one physician's behavior. Results are presented in Table A5. Consistent with medical guidance, the most sizable determinants of C-section risk are previous C-section, breech presentation, and placenta previa. Estimates are clinically consistent in direction; for example, we observe that C-section becomes more likely with age.

This model provides us with a "consensus" estimate of how different patient observables

[6] Gregory et al., 2002 and Asch, 2009 provide guidance on identifying clinical indications from ICD codes, which we follow given we are using claims data. Currie and MacLeod, 2017 and Robinson et al., 2023 use birth record data. These vary in their level of detail; we have access to some variables unavailable in birth records and can also occasionally use more granular codes, but analogously the birth records data on occasion has access to information we struggle to code from claims.

[7] We do not include race in our estimate of C-section appropriateness, but focus on medical characteristics. To the degree that Black patients experience different treatment than White patients due to medical risk factors, this should be captured in predicted appropriateness via different incidence of certain conditions.

should influence the delivery method decision. We use the parameters to estimate each patient's individual C-section risk:

$$h_i = \hat{\beta} X_i \tag{2}$$

We use this measure in two ways. First, as a control in regressions estimating residencies' C-section propensity. Second, in a measure of a physician's deviation from medical consensus. For example, if a patient has a C-section risk score of 0.2, we interpret that as 20% of physicians would perform a C-section. If the physician chooses to perform a C-section, we interpret that as a deviation from clinical consensus.

Table 1: Deviations from Predicted Delivery Method

(a) Threshold: 15%

	Predicted C-Section Risk	
Delivery Method	≤15%	≥85%
Vaginal	325,308	6,087
C-Section	56,919	54,731
N	443,045	
Deviations Total (Share)	63,006 (14.22%)	

(b) Threshold: 20%

	Predicted C-Section Risk	
Delivery Method	≤20%	≥80%
Vaginal	918,882	15,889
C-Section	198,825	123,342
N	1,256,938	
Deviations Total (Share)	214,714 (17.08%)	

(c) Threshold: 25%

	Predicted C-Section Risk	
Delivery Method	≤25%	≥75%
Vaginal	1,225,595	35,882
C-Section	304,683	233,321
N	1,799,481	
Deviations Total (Share)	340,565 (18.93%)	

Notes: Deliveries are restricted to claims in Medicaid 2015-2019, with active physicians (at least 100 deliveries over the sample period), for whom residency location is identified and where the residency has at least 10 active alumni.

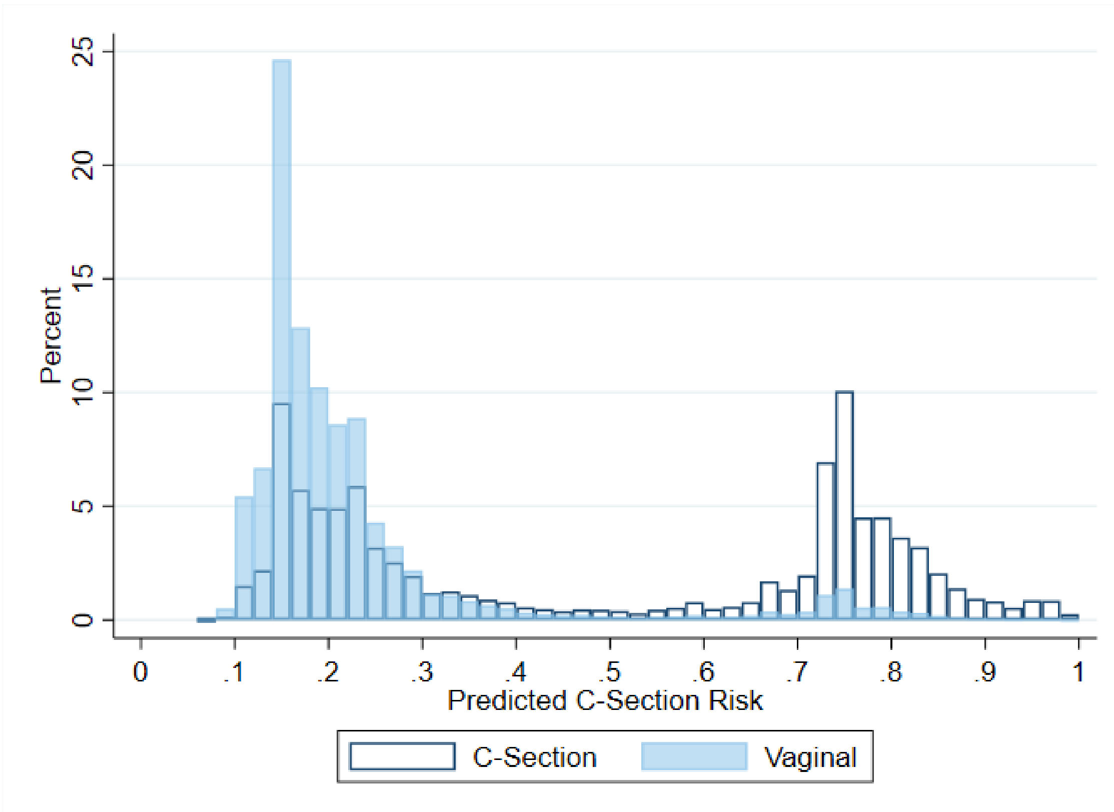
Table 1 shows the classification of the deviation variable for three different risk thresholds: 15% (i.e., below 15% risk or above 75% risk), 20%, and 25%. The share of deliveries classified as deviating from clinical consensus is between 15-19%. We use the middle threshold to create our variable indicating deviation from clinical consensus, equal to 1 if the patient is low-risk ($< 20\%$ probability) and a C-section is performed, or the patient is high-risk ($> 80\%$ probability) and a C-section is not performed.[8]

In Figure 3, we plot the distribution of C-section risk score by delivery type. The

[8] As a robustness check, we present results for an alternative measure of deviation in delivery method diagnosis. For every patient, we measure the deviation between the physician's chosen delivery method (1 for a C-section, 0 for vaginal delivery) and the predicted C-section risk, and square the result. This provides a continuous measure of the degree of disagreement between the physician's choice of delivery method and medical consensus.

distributions are concordant with evidence in Currie and MacLeod, 2017.[9] Vaginal deliveries are generally well-targeted. C-sections, however, are much less accurately targeted. Though there is a notable mass at higher risk levels, there is also a mass among low-risk patients.

Figure 3: Distribution of C-Section Risk Score by Delivery Method



Notes: Histogram plots the distribution of predicted C-section risk score by delivery method (vaginal versus C-section). Y-axis shows the percent of each group with a given risk score. C-section risk is predicted using the estimated logit model detailed in Table A5. N = 7,831,460. The sample includes all deliveries identified in Medicaid claims, 2015-2019.

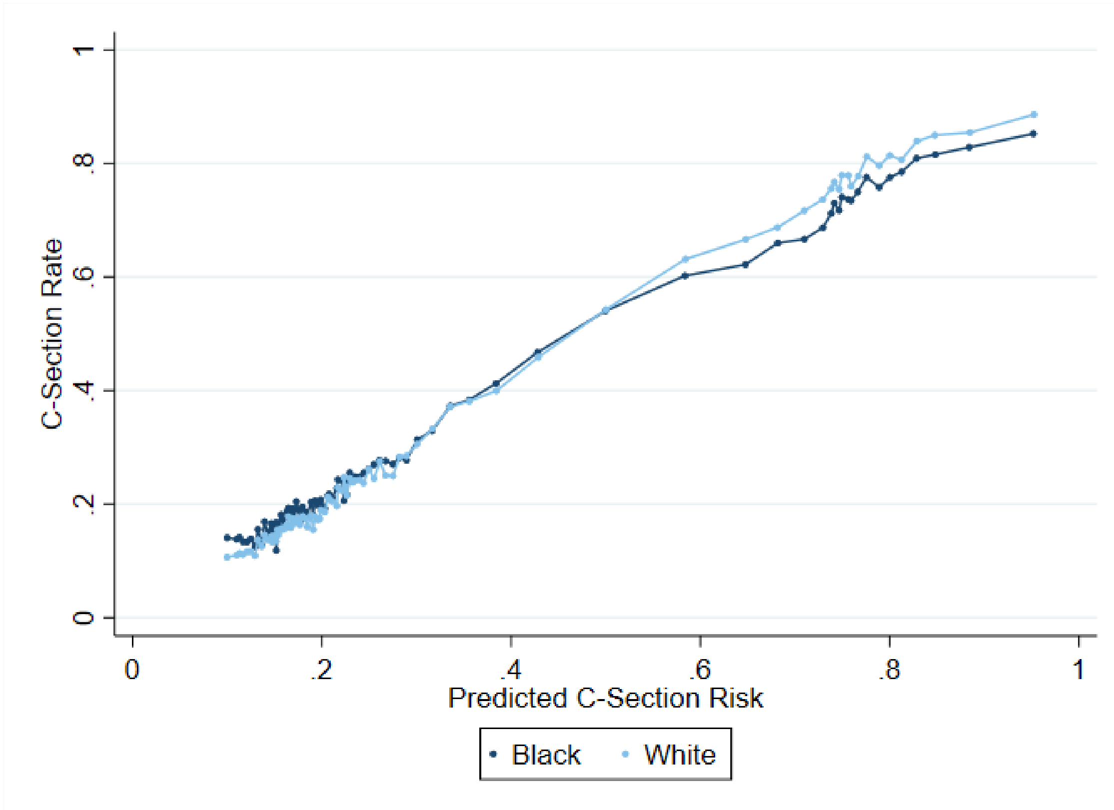
2.5 Documenting Racial Disparities in Treatment Use and Outcomes

In this section, we focus on mistargeted C-sections, examining the split by race. In Figure 4, we plot the C-section rate by percentile of predicted risk for Black and White patients separately. Among low-risk patients, a Black patient is more likely to receive a C-section than a White patient, while the reverse is true for high-risk patients. This indicates C-sections are

[9]This figure is a replication of Figure 3 in Currie and MacLeod, 2017.

less accurately targeted for Black patients, compared to their White counterparts.

Figure 4: C-Section Rates by Race and Predicted C-Section Risk

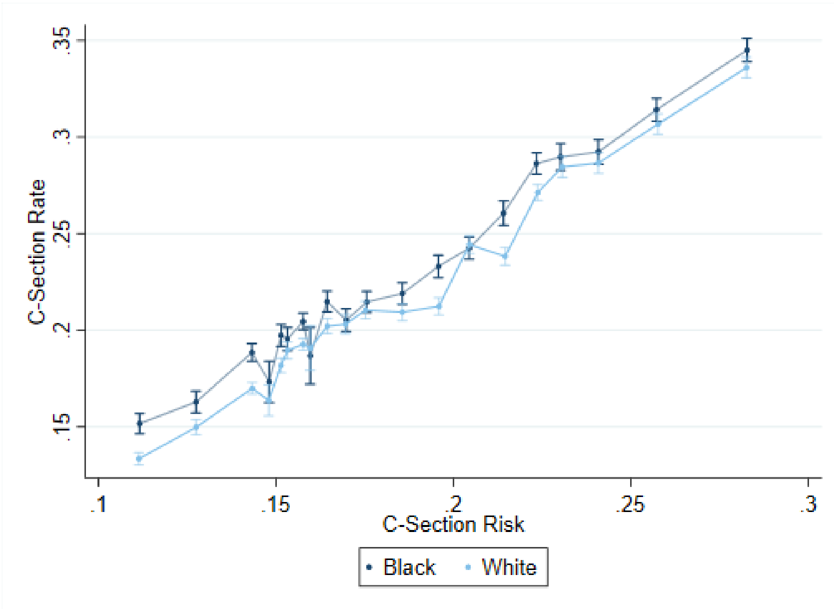


Notes: Scatter plot shows C-section rates by patient race (Black or non-Hispanic White) and predicted C-section risk. The sample is restricted to patients identified as non-Hispanic White or Black, split into 100 quantiles by predicted C-section risk, and the C-section rate is calculated within each by patient race. C-section risk is predicted using the logit model detailed in Table A5. (Note this model excludes race as a predictive factor.) N = 4,210,202.

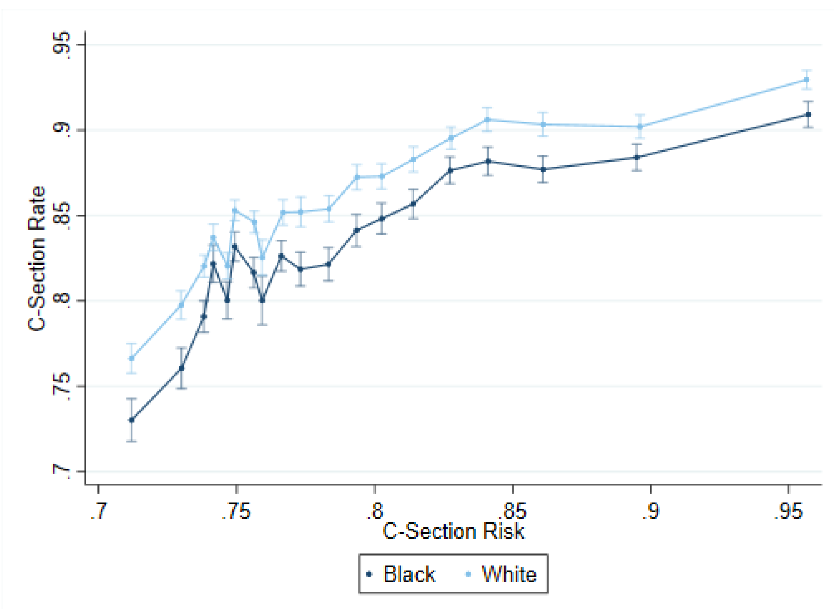
We split the sample into two groups: those with a low C-section risk (0.3 or below) and those with a high C-section risk (0.7 or above). For each, we create a scatter plot, split by race (Black or White). The result, in Figure 5, shows in greater detail that the appropriate delivery method is less accurately diagnosed for Black patients, as compared to White patients. Black patients with low risk scores are more likely than White patients with similar scores to receive a C-section, and the reverse is true for patients with high risk scores. This finding is consistent with Robinson et al., 2023, who use NVSS birth records data.

Figure 5: C-Section Rates by Race and Risk

(a) Low-Risk



(b) High-Risk



Notes: Scatter plots show C-section rates by patient race (Black or non-Hispanic White) and predicted C-section risk, calculated using an estimated logit model detailed in Table A5. (Note this model excludes race as a predictive factor.) The sample is split by predicted C-section risk into (a) patients with predicted risk $\leq 30\%$, and (b) patients with predicted risk $\geq 70\%$. Each subsample is split into 20 quantiles, and the C-section rate is calculated within each by patient race. Error bars denote 95% confidence intervals around the percentile-race C-section rate. The mean C-section rate among low-risk Black patients is 0.235 , and among low-risk White patients is 0.216 . The difference is statistically significant (t-statistic -22.798, p-value 0.000). The mean C-section rate among high-risk Black patients is 0.834 , and among high-risk White patients is 0.854 . The difference is statistically significant (t-statistic 14.433, p-value 0.000). $N_{\text{low risk}}$ = 1,156,817, and $N_{\text{high risk}}$ = 300,808.

In Table 2, we detail the rates of delivery complications in our sample, broken out by delivery method and race. Around 9.4% of deliveries experience any complication. The most common complications are those related to hemorrhage (in both delivery types), occurring in about 4.5% of deliveries. All complications are more prevalent in Black patients than White patients, except for laceration, though the differences vary in magnitude. Overall, complications are approximately 10% higher among Black patients versus White patients.

Table 2: Summary Statistics for Delivery Complications by Race

	Mean (Std. Dev.)			Black-White Δ
	All	Black	White	
Any Complication	9.39 (29.17)	9.81 (29.75)	8.87 (28.44)	0.94***
Vaginal Delivery				
Laceration	3.31 (17.89)	3.18 (17.56)	3.26 (17.77)	-0.08*
Hemorrhage	4.42 (20.56)	4.76 (21.29)	3.90 (19.35)	0.86***
Thrombotic/Other	3.36 (18.01)	3.49 (18.36)	3.35 (17.99)	0.14***
C-Section				
Hemorrhage	3.62 (18.68)	4.19 (20.03)	3.02 (17.11)	1.17***
Infection	0.13 (3.55)	0.14 (3.77)	0.11 (3.38)	0.03***
Operative	0.13 (3.66)	0.18 (4.21)	0.11 (3.27)	0.07***
Thrombotic/Other	3.91 (19.38)	4.12 (19.87)	3.91 (19.39)	0.21***
N	3,098,437	611,349	1,008,412	

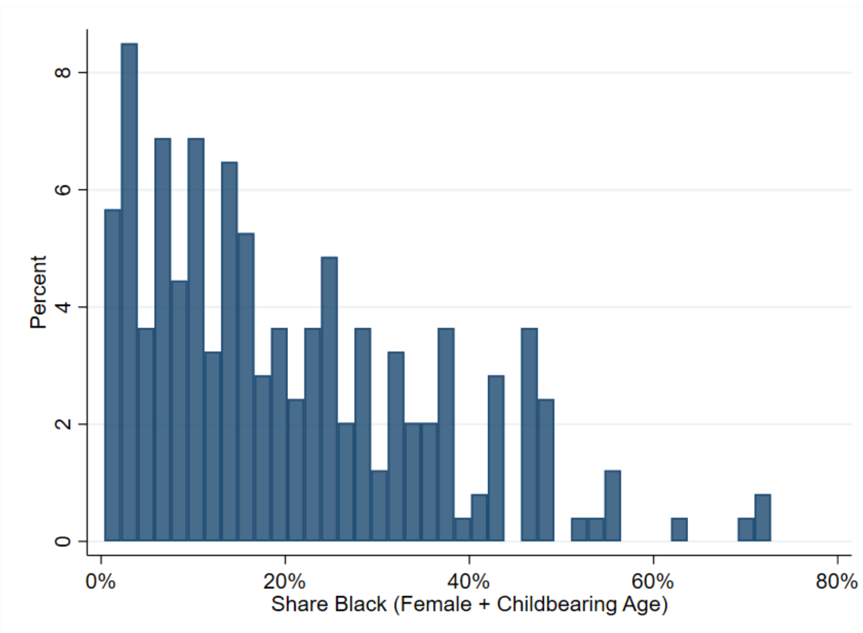
Notes: Table details summary statistics for delivery complications by delivery mode and race. An observation is a delivery (patient-physician pair linked to a birth outcome). Deliveries are restricted to claims in Medicaid 2015-2019, with active physicians (at least 100 deliveries over the sample period), for whom residency location is identified. Column 4 reports the difference in means and significance from a two-tailed t-test, * indicates $p < 0.05$, ** for $p < 0.01$, and *** for $p < .001$.

2.6 Racial Patient Mix at Residency

One hypothesis we explore is that training with a more diverse patient mix reduces the gap in outcomes between patients of different races. To test that conjecture, we measure a residency's patient mix based on demographic data from the ACS. In Figure 6 we plot the distribution of the Black share of the female and childbearing age population across residencies. This share is computed at the 3-digit ZIP code level. Though the distribution is skewed, the range is large. The mean share of Black females in the childbearing age population in a residency ZIP-3 is 20.48%, the median is 15.96%, with a standard deviation of 15.57%. The 5^{th} percentile is 1.86% and the 95^{th} percentile is 47.60%.

This shows there is substantial variation in the patient mix of a residency's locality, meaning physicians may leave training with highly variable experience diagnosing and treating conditions that differ across demographic groups. As an example, consider the University of Michigan in Ann Arbor, versus the University of Pennsylvania in Philadelphia. These residencies are typically both ranked in the top twenty in the US. But the share of the population of Ann Arbor that is Black is approximately 7%, while in Philadelphia it is 40%.

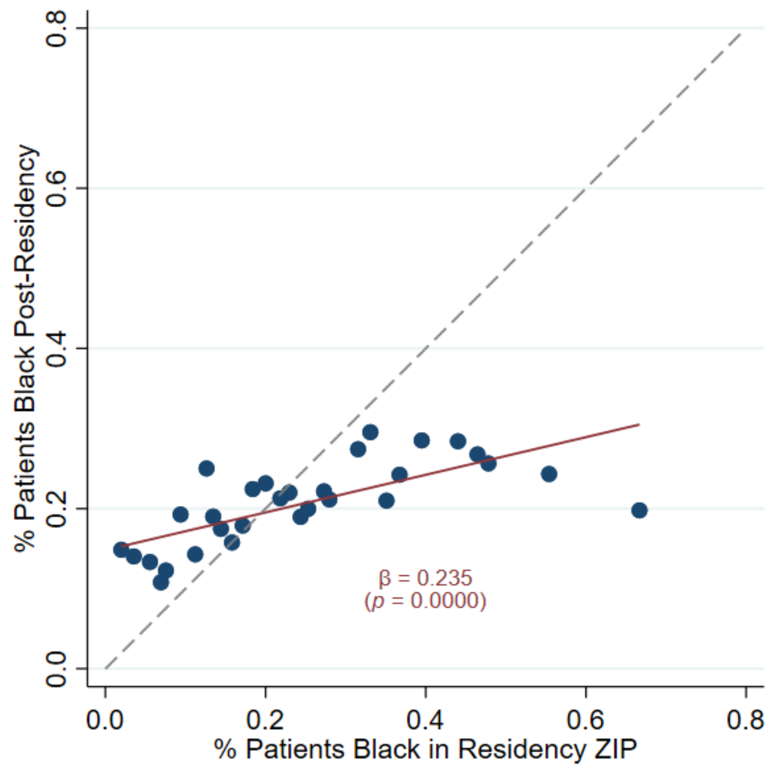
Figure 6: Distribution of Share of Population Black (Female and Childbearing Age) Across Residency ZIPs



Notes: Histogram shows the distribution of the Black share of females of childbearing age within a residency's 3-digit ZIP code. Data comes from the American Community Survey. The mean share is 20.48%, median is 15.96%, and standard deviation is 15.57%. The 5^{th} percentile is 1.86% and the 95^{th} percentile is 47.60%. $N = 247$ residencies.

We next examine the relationship between residency and post-residency characteristics. Figure 7 plots physicians' racial patient mix post-residency (in Medicaid claims) versus the racial patient mix of their residency. There appears to be a positive but slight relationship between the two, suggesting that there is not necessarily path dependence in patient mix between residency training and post-residency practice. An important caveat is that, because we use Medicaid claims data, we observe a large but incomplete fraction of deliveries, and therefore an incomplete picture of physicians' post-residency practice. In particular, because Medicaid's population skews less White than the US population as a whole, the share of a physician's patients that are covered by Medicaid (which we cannot observe) will likely mediate post-residency patient mix.

Figure 7: Binned Scatter of Racial Patient Mix at Residency versus in Post-Residency Practice



Notes: Binned scatter plot shows racial patient mix at residency versus racial patient mix in post-residency practice. Racial patient mix calculated as the Black share of (i) female population of child-bearing age in residency ZIP, or of (ii) post-residency claims in Medicaid 2015-2019 associated with a delivery. We divide the sample into 30 quantiles. The coefficient from a regression of practice patient mix on residency patient mix is displayed with the associated p-value. Both the binned scatter and regression are weighted by the number of deliveries a physician performs. Data restricted to active physicians (at least 100 deliveries over the sample period), for whom residency and residency patient mix is identified from the AMA MasterFile and ACS data. The 45° line is plotted in dashed grey. $N = 7{,}679$ physicians.

3. Empirical Strategy

3.1 Identification

The ideal experiment to assess the impact of specific residencies on physician performance for patients of different races would be to first randomly allocate physicians to residency programs. Then, to assess the effect of their training, ideally, we would randomly allocate patients to different physicians in post-residency practice. In practice, this is clearly not possible, and selection is a concern at both stages.

In this paper, we abstract from selection into residencies and aim to understand the degree of variation across residency alumni in post-residency treatment use and outcomes, being transparent that some of this could be driven by selection by physicians into different programs. One concern, for example, would be that higher-quality physicians select into higher-quality residencies. It is worth noting, however, that some residency characteristics of interest—such as racial patient mix—do not appear to be correlated with the typical dimensions of selection we would be concerned with, such as quality. An anecdotal example of this might be Johns Hopkins Hospital in Baltimore, Maryland, and the Mayo Clinic in Rochester, Minnesota, ranked second and third for OB/GYN work, respectively. The former is located in a city where the Black share of the population is above 60%, while the latter is located in a city where the Black share of the population is below 10%. The second concern is selection and matching between patients and physicians. For example, physicians who trained at higher-quality residencies might be systematically paired with more complex patients, who have worse outcomes. Physicians who are passionate about racial equity in healthcare might attend more diverse residencies and make efforts to pair with Black patients in their post-residency practice.

Several characteristics of delivery, however, allow us to mimic quasi-random allocation of patients to physicians who trained at different residencies. We focus on a large subset of deliveries with quasi-random variation in patient-physician matching driven by the spontaneous onset of labor and rotational variation in physicians' inpatient shifts. This technique mirrors strategies used in the "judge leniency" literature (Doyle et al., 2015; Eichmeyer and

Zhang, 2022; Kling, 2006). The emergent nature of the deliveries we consider, and rotational variation in physician shifts, in effect randomly assigns patients to physicians with different residency training on a given labor and delivery ward.

The assignment is quasi-random because of two factors. First, labor is spontaneous (emergent) for most patients. We identify patients who enter labor spontaneously using diagnosis codes through prenatal care and delivery. These codes tell us if a patient's labor was planned in advance, and we discard such deliveries from our sample. Second, physicians rotate who is on call at a delivery ward throughout each week. Such shifts are typically decided at least one month in advance. Together, these two factors quasi-randomly assign patients to physicians within a hospital.[10,11] If physicians within a hospital vary in which residency they attended, this will also be quasi-randomly assigned as part of that process. We use this to estimate variation across residencies in C-section usage and patient outcomes.

To check our assumption of quasi-random assignment, within a hospital, we regress our residency characteristics of interest (residency ranking and residency patient mix) on patient characteristics.

$$\text{Residency Characteristic}_{ijbt} = \pi_0 + \pi_1 P_i + \pi_2 X_i + \gamma_{bt} + e_{ijbt} \tag{3}$$

If patient-physician assignment is quasi-random, conditional on patient controls (X_i) and hospital-month-year of birth (γ_{bt}), then patient demographics and risk factors (P_i) should have no significant relationship with the residency characteristics of the delivery physician j. In other words, π_1 should be ≈ 0. It would be a problem, for example, if Black patients are systematically assigned to a physician who attended a residency with a higher share of Black

[10] Violations of the randomness in our assignment mechanism might occur if some physicians are systematically more likely to take certain shifts (e.g., the weekend) and a patient type is systematically more likely to go into labor at those times. There is no biological logic that would give rise to this pattern. Another concern may be if physicians organize shifts around their patients due dates. This would be challenging for physicians to satisfy for all patients, not least given due dates are an imprecise prediction of birth timing. The most commonly cited study we found, a retrospective analysis of nearly 20,000 patients in Australia, found that approximately 5% of births occurred on the estimated due date (Khambalia et al., 2013).

[11] We note that Medicaid claims do not explicitly attach a claim to a hospital, but do identify a billing provider as the entity responsible for billing a patient for healthcare services, and categorize the billing provider's type (e.g., hospital, birth center). Billing is typically performed by the hospital, and covers charges for the delivering provider as well as the facilities used. We restrict our sample to deliveries with a hospital as the billing provider, so that we can identify the facility at which the delivery is performed.

patients. Or, if a complex patient is assigned to a physician who attended a higher-ranked residency.

Results for this test of random assignment are displayed in Figure 8. These demonstrate that while patient characteristics can predict our clinical outcomes of interest (delivery complication and C-section), they have no discernible impact on the assigned physician's residency rank, nor experience with different patient groups. The F-statistic for joint significance in both regressions on physicians' residency characteristics is below 2.5; for the regressions on clinical outcomes, it is two orders of magnitude larger, at least. Notably, patient race is not associated with a physician's patient mix at training.

We leverage this quasi-random assignment of patients in labor to physicians, within the hospital, to estimate fixed effects for residencies for C-section delivery and delivery complications, both overall and split by race (White and Black). To improve precision, we restrict to residencies with at least 10 physicians practicing in Medicaid between 2015-2019.[12] This leaves us with a sample of 2,861,969 deliveries to 9,633 physicians from 272 residencies. Additionally, in our regression on C-section delivery, we restrict to nulliparous patients only (as the delivery method of higher order births is almost entirely determined by prior delivery method).

For each outcome Y for patient i delivering at hospital b with physician j, who attended residency r, in month-year t, we run the following regression:

$$Y_{ijrbt} = \alpha + \gamma_{r(j)} + \gamma_{bt} + \beta X_{it} + \varepsilon_{ijrbt} \tag{4}$$

Where γ denotes fixed effects for residency (subscript r for physician j) and hospital-month-year (subscript bt, to reflect the assignment mechanism). X_{it} is a vector of patient controls. For the delivery method, we control for our measure of C-section risk. For delivery complications we include controls for nulliparity, singleton, term and vertex, as well as 21 comorbidities

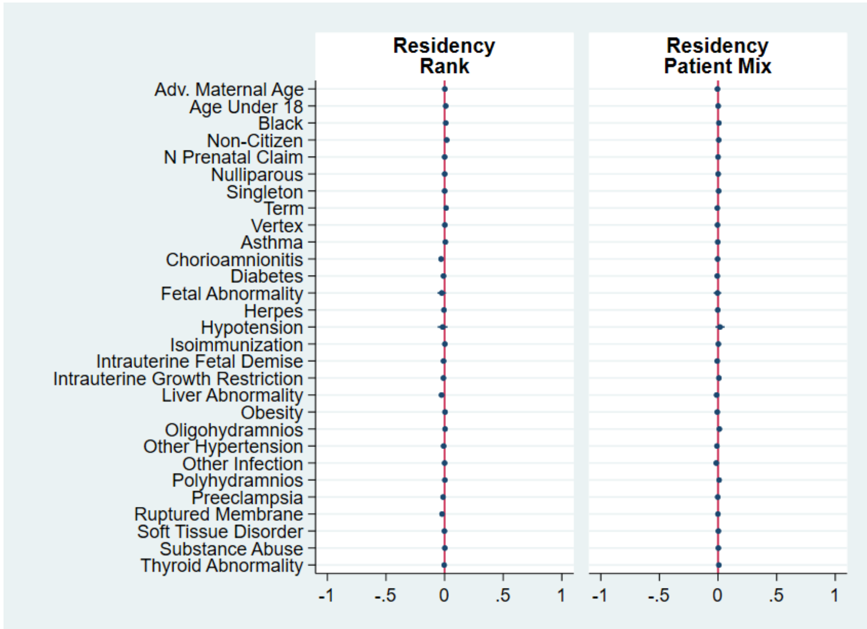
[12] We provide additional sample descriptives on the distribution of (a) the number of residency alumni, and (b) the number of deliveries per residency in Appendix B.

chosen based on clinical guidance available in Asch, 2009.[13] Standard errors are clustered at the physician level.

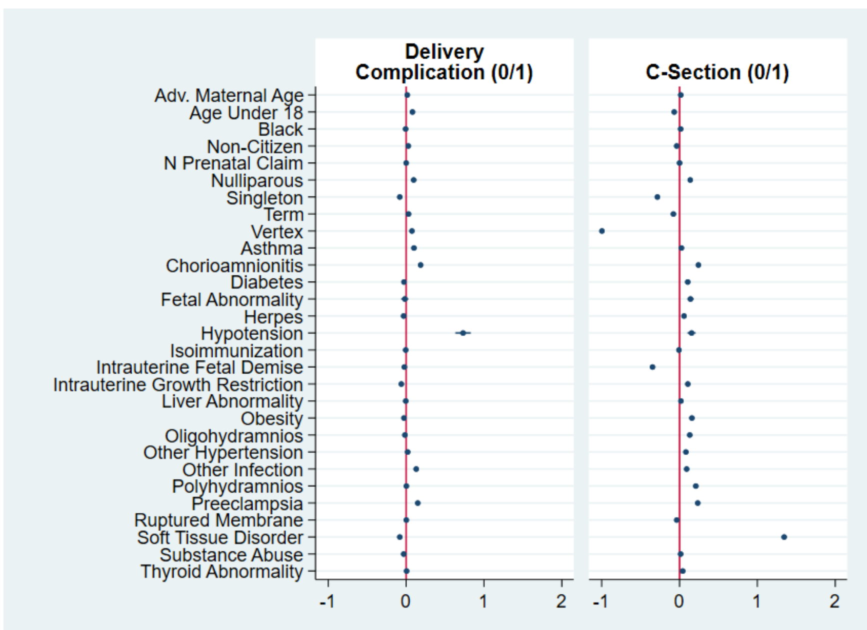
[13]These include: asthma, chorioamnionitis, diabetes, fetal abnormalities, herpes, hypotension, isoimmunization (Rh sensitivity), intrauterine fetal demise, intrauterine growth restrictions, liver abnormalities, macrosomia, obesity, oligohydramnios, hypertension, infection, polyhydramnios, preeclampsia, ruptured membranes, soft tissue disorders, substance abuse, and thyroid abnormalities.

Figure 8: Association of Patient Characteristics with (a) Assigned Physician's Residency Characteristics and (b) Clinical Outcomes

(a) Assigned Physician's Residency Characteristics



(b) Clinical Outcomes



Notes: Plots show the coefficients and confidence intervals (at 95% level) for regressions of four variables on patient characteristics. Dependent variables include: physician's residency rank, physician's residency patient mix, an indicator for delivery complication, and an indicator for C-section. All dependent variables are standardized. Controls include hospital-month-year fixed effects and standard errors are clustered at the physician level. Sample restricted to spontaneous deliveries in hospitals in Medicaid claims, 2015-2019. Physicians restricted to those with observable residency patient mix, and at least 100 deliveries over 2015-2019. The F-statistic for the regression on: residency rank is 2.16 , residency patient mix is 1.54 , delivery complication is 168.82 , and C-section is 2356.91 .

An important source of variation for our empirical strategy is physician mobility post-residency, such that physicians from multiple different residencies end up practicing in the same location. In Figure 9 and Figure 10, we show maps of obstetricians' and gynecologists' residency location by state of practice for California, Texas, Pennsylvania, and Florida. All the maps demonstrate substantial migration after residency, both across and within states.

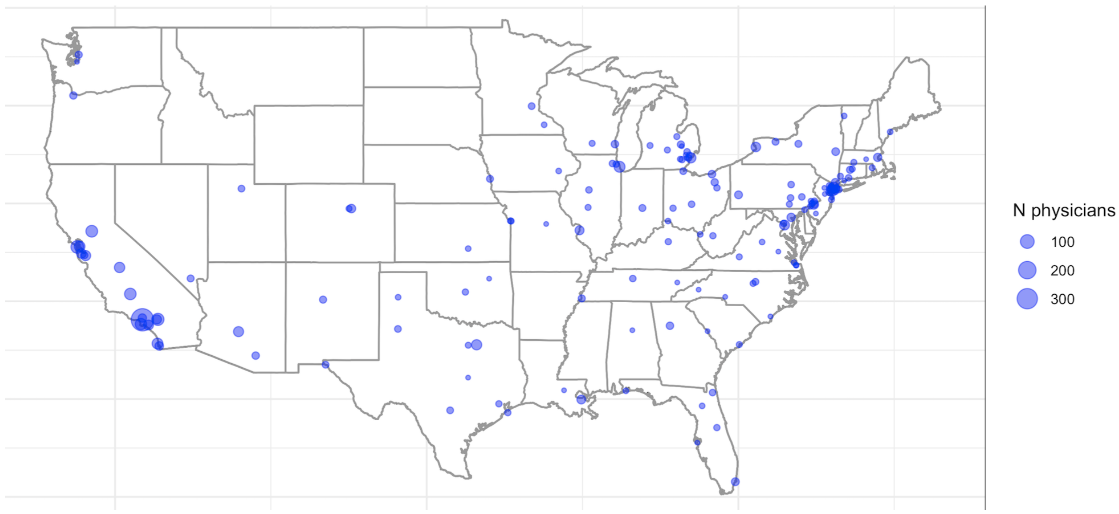
3.2 Interpretation

A residency's fixed effect captures how physicians who trained in that residency tend to perform relative to physicians from other residencies, after accounting for differences in patient characteristics, hospital, and time. For the regression on delivery method (an indicator for C-section delivery), the effect measures the strength of a residency-specific practice style in delivery type (or C-section intensity). For the regression on delivery complications (an indicator for experiencing any complication), the effect measures the marginal probability of a patient experiencing a complication if she sees a physician who attended that residency. The fixed effects can be interpreted as answering: "Do graduates of this particular program, on average, deliver different maternal outcomes than the typical physician trained elsewhere?"

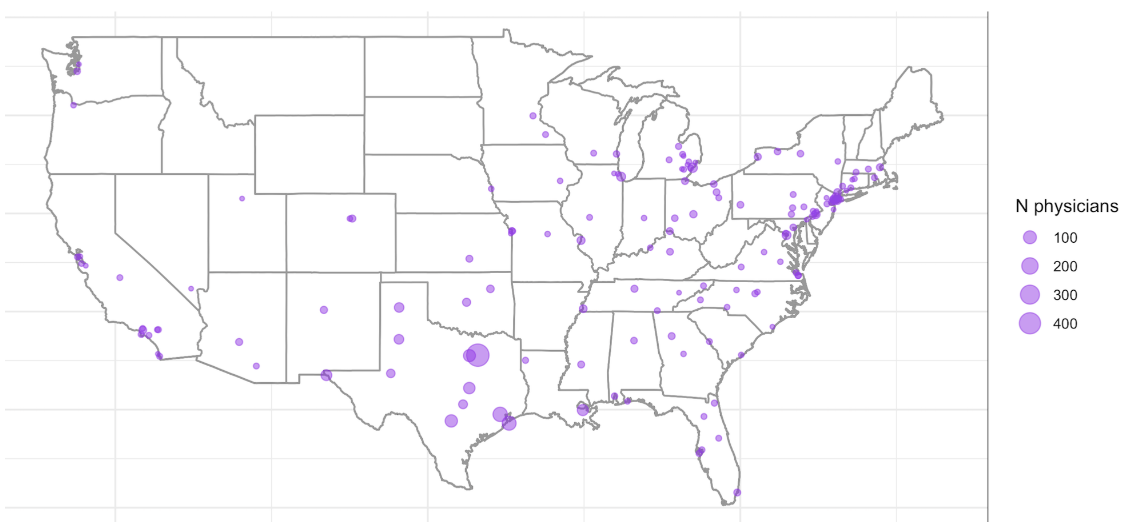
Note that our residency fixed effect captures the effect on the patient of being treated by a physician who attended a particular residency, *not* the effect of that residency on the patient. The distinction is important given the potential selection of physicians into residencies. For example, one reason we might estimate a higher C-section intensity for a residency is that physicians who have a preference for surgery select into that residency.The estimated residency fixed effects could reflect multiple mechanisms by which physician training impacts post-residency practice style. One of those could be physician selection into training institutions with like-minded peers. Another could be a distinct "style" of practice that residencies impart onto their physicians. In conversations with obstetricians, several such "training signatures" were mentioned for particular universities, with some being known for higher C-section intensity.

Figure 9: OB/GYN Residency Location by State of Practice: California and Texas

(a) California



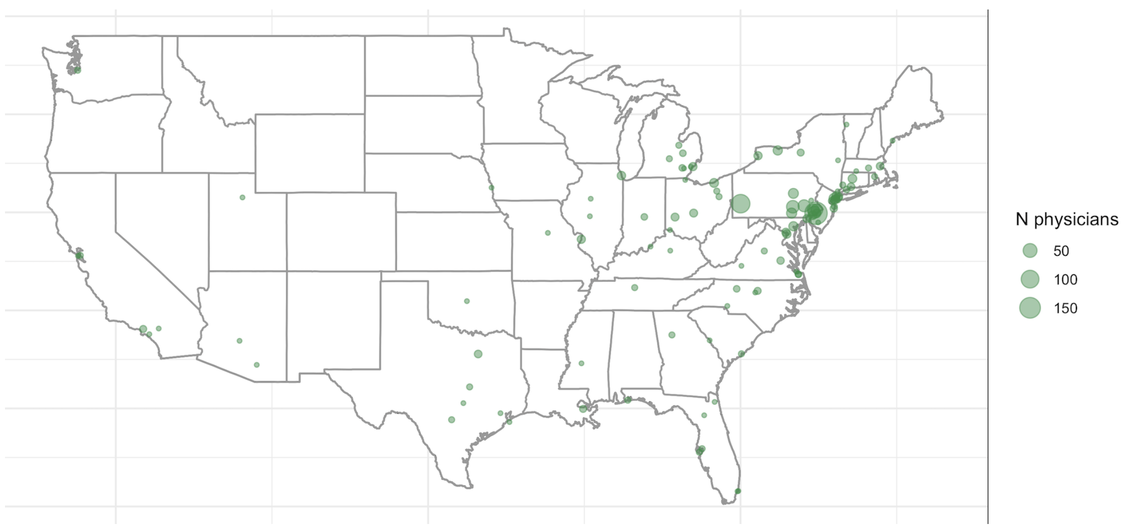
(b) Texas



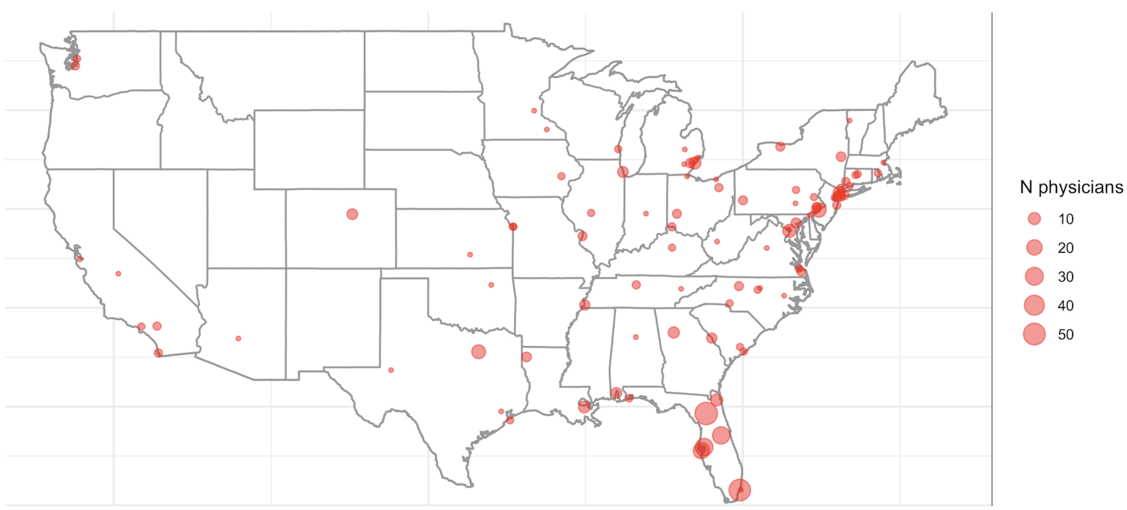
Notes: Maps show the residency locations for physicians practicing in (a) California or (b) Texas according to Medicaid claims, 2015-2019. Circle sizes indicate the number of physicians who trained in a particular residency location.

Figure 10: OB/GYN Residency Location by State of Practice: Pennsylvania and Florida

(a) Pennsylvania



(b) Florida



Notes: Maps show the residency locations for physicians practicing in (a) Pennsylvania or (b) Florida according to Medicaid claims, 2015-2019. Circle sizes indicate the number of physicians who trained in a particular residency location.

4. Results

In this section we outline our results on the variation across residencies in physicians' post-residency practice for patients of different races, and test the hypothesis that a more diverse patient mix at residency improves diagnostic ability in treatment use.

4.1 Do Residencies Vary in Physician Practice Style and Performance by Patient Race?

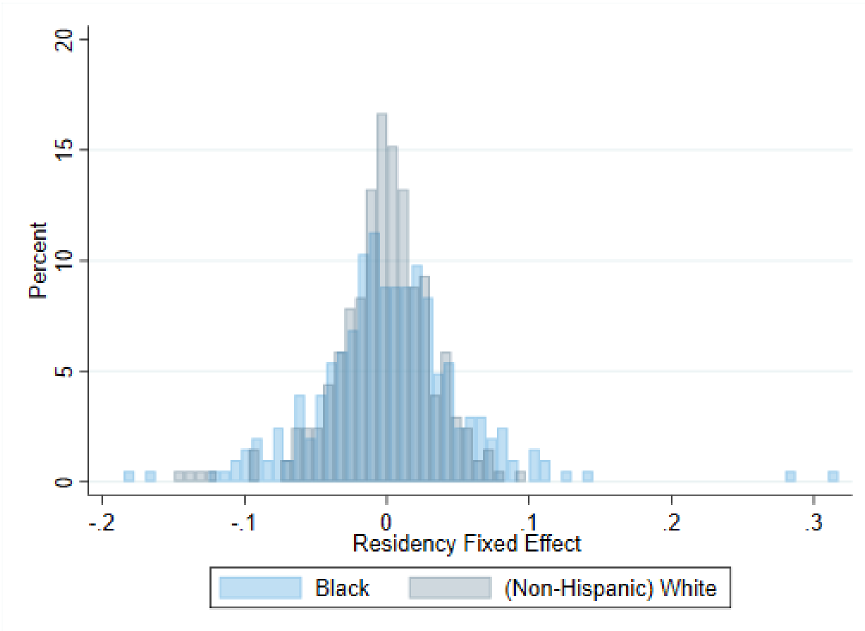
We first examine the relationship between the ranking of residencies across the two outcomes by race. For example, does a residency that ranks better for its effect on complications for White patients also rank better for its effect on complications for Black patients? Do physicians from the same residency who perform relatively more C-sections on Black patients also perform relatively more C-sections on White patients?

We assess these questions by comparing the fixed effects of each residency across the regressions, which are split by race. A residency's fixed effect on C-sections can be interpreted as the relative C-section intensity of that residency's alumni, controlling for hospital, time, and individual patients' C-section risk. In other words, the relative effect on the probability of receiving a C-section when assigned to an OB/GYN from Residency A, compared to an OB/GYN from Residency B, within the same hospital-month-year. For each residency, this fixed effect is estimated both overall and separately for Black versus White patients. Similarly, a residency's fixed effect on delivery complications is the relative effect of seeing physicians from that residency on a patient's probability of experiencing a complication, again within hospital-month-year and controlling for over twenty clinically informed comorbidities. This is also estimated separately for Black versus White patients.

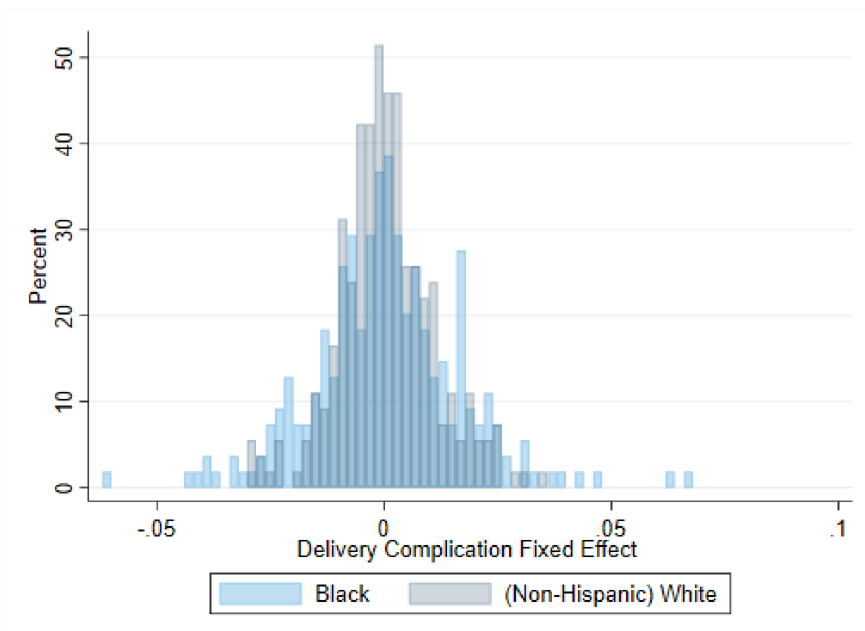
In Figure 11 we plot the distribution of residency fixed effects by race for C-section delivery (Figure 11a) and delivery complications (Figure 11b). The mean C-section rate for Black (White) patients in the sample is 0.33 (0.31), with a standard deviation 0.47 (0.46). The mean delivery complication rate for Black (White) patients in the sample is 0.10 (0.09), with standard deviation 0.31 (0.29).

Figure 11: Distribution of Residency Fixed Effects, by Patient Race

(a) Residency Fixed Effect on C-Sections



(b) Residency Fixed Effect on Delivery Complications



Note: Histograms plot the distribution of residency fixed effects from Equation (4), estimated separately for Black and non-Hispanic White patients. The dependent variable in Panel (a) is an indicator for C-section delivery, and in Panel (b) is an indicator for delivery complications. Controls include: assignment generating quasi-random physician-patient matching (hospital-month-year fixed effects) and (a) C-section risk or (b) patient comorbidities informed by Asch, 2009. Deliveries are restricted to claims in Medicaid 2015-2019, with active physicians (at least 100 deliveries over sample period), for whom residency location is identified. The number of residencies is 272. The mean residency fixed effect for C-sections for Black (White) patients is 0.0019 (-0.0019). The mean residency fixed effect for delivery complications for Black (White) patients is 0.0004 (0.0003).

We begin with a simple assessment of the overlap for between the top and bottom quartiles of the fixed effects for delivery method and complications, by race. We rank residencies by their C-section fixed effect on Black and White patients separately, and label those in the top (more intense) versus bottom (less intense) quartiles. We do the same for residencies on their complication fixed effect; those in the top quartile suppress the complication rate more than those in the bottom quartile.

The results are displayed in Table 3. One might expect that the quality of physician that a residency produces is relatively stable across patient races. Our results suggest that, though positively related, a residency's relative effect on C-section use among patients of one race can differ substantially from its relative effect on C-section use among patients of another race. The same is true for the fixed effects on delivery complications.

Table 3: Quartiles for Residency C-Section and Delivery Complication Fixed Effects, by Patient Race

(a) C-Section Fixed Effects

Black \ White	Top 25%	Bottom 25%
Top 25%	24	18
Bottom 25%	10	24

(b) Delivery Complication Fixed Effects

Black \ White	Top 25%	Bottom 25%
Top 25%	26	10
Bottom 25%	15	26

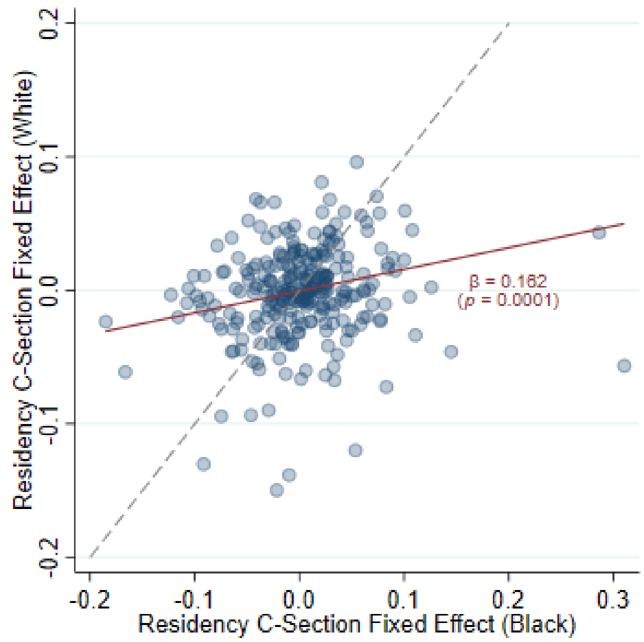
Notes: Table details the overlap between residencies in the top and bottom quartiles of fixed effect for (a) C-section delivery and (b) delivery complications, by patient race (Black and White). The fixed effects are estimated from two regressions of the form Equation (4), one for Black and one for White patients. Controls include: assignment generating quasi-random physician-patient matching (hospital-month-year fixed effects) and (a) C-section risk or (b) patient comorbidities informed by Asch, 2009. Deliveries are restricted to claims in Medicaid 2015-2019, with active physicians (at least 100 deliveries over the sample period), for whom residency location is identified. The number of residencies is 272.

We examine this in more detail with scatter plots in Figure 12 charting the fixed effects of residencies on C-section rates and delivery complications for Black versus White patients. We include the coefficient and significance from a simple linear regression between the fixed effects for the same measure, but different patient races. In both cases the fixed effects have a positive but only very slight relationship, with coefficients of between 0.1-0.2: suggesting a one percentage point increase in (for example) a residency's C-section fixed effect for Black patients translates to a 0.1 percentage point increase in the same residency's C-section fixed effect for White patients.

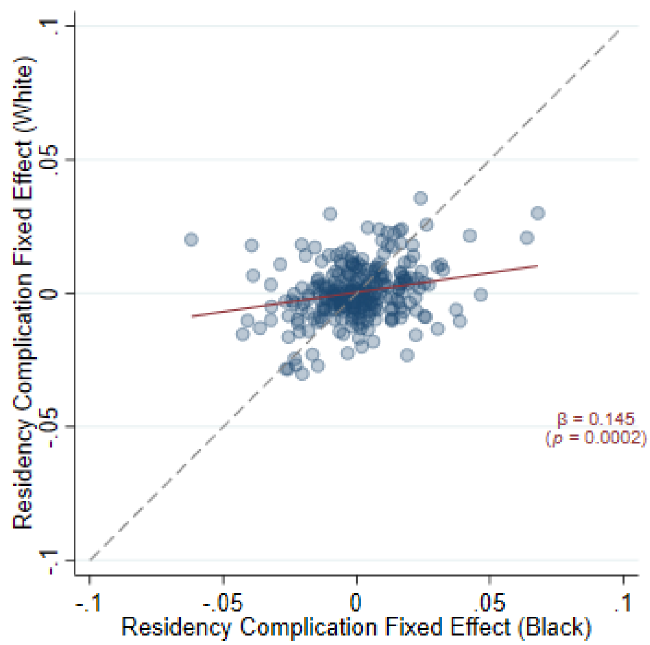
These mismatches suggest non-uniform learning (or application of learning) across patients of different races. This could come from a few characteristics of residency. Training may have differential emphasis on the clinical needs of different patient groups. Bias within training—from instructors or within clinical settings—may also impact how residents learn to treat patients from different racial backgrounds. Another potential explanation is lack of exposure: residencies located in less diverse areas may not provide sufficient exposure to patients of different racial backgrounds. This could limit the ability of physicians to generalize their training across groups. We explore this hypothesis next.

Figure 12: Scatter of Residency Fixed Effects for Black and White Patients

(a) C-Section



(b) Delivery Complications



Notes: Scatter plot of a residency's fixed effect on (a) C-section rates and (b) delivery complications, estimated from two regressions of the form Equation (4), one for Black and one for White patients. Controls include assignment fixed effects and (a) C-section risk or (b) patient comorbidities. Deliveries are restricted to claims in Medicaid 2015-2019, with active physicians (at least 100 deliveries over the sample period), for whom residency location is identified. The number of residencies is 272. A regression coefficient (β) of the White fixed effect on the Black fixed effect is displayed, as well as its p-value.

As a test of the significance of the estimated fixed effects, we take a permutation approach. First, we construct a set of summary measures of the estimated fixed effects as test statistics. Those are the mean, median, IQR, and the sum of squared fixed effects (SSF $= \sum \gamma_r^2$), to capture how extreme the estimated values are, for both outcomes (C-section and delivery complication) and all demographic groups (all, White, and Black). We also calculate the share of residencies in different quartiles by race; for example, in the top 25% of residency fixed effects for C-sections on White patients, but the bottom quartile for C-sections on Black patients.

We then randomly reassign residency labels across deliveries in our sample, keeping the number of deliveries per residency the same, and re-estimate the residency fixed effects and associated statistics. We repeat this step 500 times to generate an empirical distribution for every statistic, under the null that residency is effectively driven by random variation (preserving the sample error driven by different numbers of observations per residency). We then compare our actual estimates to the empirical distribution for each, as a measure of the degree to which residencies generate non-random variation in practice style.

Our sample statistics estimated from the fixed effects are displayed in Table 4. The p-values from a two-tailed test for difference from the empirical distribution under random residency are shown in parentheses. Almost all measures are statistically significant, suggesting that our results differ from the case where residency fixed effects reflect essentially random variation. In particular, our estimates of the share of residencies in non-overlapping quartiles by race, for each outcome measure, are both statistically significant at the 5% level, indicative of real differences in residencies' impact on patients of different races.

Table 4: Summary Statistics on Residency Fixed Effects by Outcome and Patient Race

	C-Section (0/1)			Delivery Complication (0/1)		
	All	White	Black	All	White	Black
Mean	-0.002	-0.002	0.002	0.000	0.000	0.000
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Median	-0.001	-0.001	0.001	0.000	-0.000	0.000
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
IQR	0.026	0.038	0.053	0.010	0.012	0.018
	(0.028)*	(0.164)	(0.192)	(0.000)***	(0.052)	(0.088)
$\sum \gamma_r^2$	0.142	0.328	0.798	0.020	0.031	0.074
	(0.004)**	(0.000)***	(0.000)***	(0.008)**	(0.000)***	(0.000)***
% residencies in non-overlapping		0.103			0.092	
quartiles by race		(0.036)*			(0.000)***	

Notes: Table details summary statistics for residency fixed effects by outcome and patient race. p-values in parentheses are calculated based on an empirical distribution for each statistic generated from 500 replications of random permutation of residency across deliveries. Deliveries are restricted to claims in Medicaid 2015-2019, with active physicians (at least 100 deliveries over the sample period), for whom residency location is identified. Included residencies must also have at least 10 alumni actively practicing in Medicaid, 2015-2019. * indicates $p < 0.05$, ** for $p < 0.01$, and *** for $p < .001$.

4.2 Does Racial Patient Mix at Residency Impact a Physician's Diagnostic Performance?

A specific hypothesis of interest is whether training with a more diverse patient mix reduces the gap in treatment choice or complications between a physician's White and Black patients. We investigate this using as a "treatment" variable the Black share of the female population of childbearing age in a residency's 3-digit ZIP area.[14] As in the regressions above, we exploit quasi-random assignment of patients to physicians by removing the residency fixed effect and adding a control for residency patient mix.

In effect, our empirical strategy compares the treatment choices for physicians who attended residencies with different patient mixes, controlling for patient-physician selection and matching. To build intuition, consider the following thought experiment. Two physicians graduate from medical school with similar performance, intending to become OB/GYNs. Through the National Resident Matching Program (NRMP), they are allocated to the University of Pennsylvania and the University of Michigan, Ann Arbor, both ranked in the top twenty OB/GYN residencies in the US. The two residencies, though similar on many dimensions, differ in the racial patient mix: Philadelphia is 42% Black, while Ann Arbor is 7% Black. After qualifying as OB/GYNs, the two physicians move to Stanford Healthcare. In the same environment, will Black patients' treatment differ across the two physicians, given the different levels of diversity in their training experience?

The specification is as follows:

$$\begin{aligned}\text{Deviation}_{ijbt} = &\alpha_0 + \beta_0 \text{Black}_i + \beta_1 \text{ResidencyPatientMix}_j \\ &+ \beta_2 (\text{Black}_i \times \text{ResidencyPatientMix}_j) \\ &+ \gamma_{bt} + \varepsilon_{ijbt}\end{aligned} \tag{5}$$

Where Deviation_{it} is a 0/1 variable equal to one if patient i's delivery method performed by physician j at hospital b in month-year t, deviated substantially from the medical consensus

[14]This ZIP-level measure has been used to measure the racial composition of a physician's patients in other work, such as Alsan et al., 2024.

predicted by our logit model. We define a substantial deviation as one where fewer than 20% of physicians would have performed the same delivery method on patient i as physician j. Fixed effects control for the assignment of patient to physician, as well as seasonality in births. Standard errors are clustered at the physician level. Black_i is an indicator equal to one if the patient is Black, zero otherwise. We focus on the Black-White race gap, and so restrict to patients who identify as either Black or White.

The coefficient β_0 captures the differential rates of deviation in Black patients' delivery method diagnoses (relative to non-Hispanic White patients). In particular, we analyze how that differential deviation changes when we include patient mix at physician residency as a covariate ($\text{ResidencyPatientMix}_j$), and the interaction of the two. If a more diverse patient mix at residency improves a physician's ability to treat patients of different races, we would expect that the coefficient on the interaction would be small and/or negative; in other words, it suppresses any error in delivery method diagnosis.

There are two potential mechanisms by which training with a diverse patient mix could improve a physician's patient outcomes. First, physicians become better able to account for clinical heterogeneity (as outlined above) through "learning-by-doing." In other words, through treating a more heterogeneous patient population. Second, physicians may improve their soft skills (or "bedside manner") and learn to adjust these appropriately given different population concerns through the same process. This may improve their ability to generate trust and improve communication with historically marginalized patients.

A more sensitive concern is that a physician may have some racial bias that drives disparities among their patients. Indeed, clinical researchers have on occasion concluded that the lack of observable health differences across patients of different races, who nonetheless receive different treatments, points to "unconscious bias and structural racism" in medical decision-making (Main et al., 2023). Suggestions have included standardized protocols to systematize process for all patients and limit physician discretion (Hamm et al., 2020), but standardization is possible only in stylized medical environments. Training with a diverse patient mix could mediate this channel. Research has suggested that exposure to different demographic groups can reduce bias–also known as the "contact hypothesis" (Alrababah et al., 2021; Bursztyn et al., 2024).

Our results, shown in Table 5, show that Black patients are more likely to be misdiagnosed in delivery method. But the training experience of their physician with a more diverse patient mix has little discernible impact on that diagnostic error. Though in sign it appears to suppress the error in direction, it is statistically insignificant.

Table 5: The Impact of Residency Patient Mix on Diagnosis Deviations by Patient Race

	Clinical Deviation	Clinical Deviation	Clinical Deviation
Black	0.0040***	0.0040***	0.0042***
	(0.0007)	(0.0007)	(0.0012)
Residency		0.0047*	0.0050*
Patient Mix		(0.0024)	(0.0027)
Black ×			-0.0008
Residency Patient Mix			(0.0038)
Constant	0.0419***	0.0408***	0.0407***
	(0.0003)	(0.0007)	(0.0007)
Mean (All)	0.0433	0.0433	0.0433
Mean (White)	0.0411	0.0411	0.0411
# NPIs (Clusters)	6,632	6,632	6,632
N	723,450	723,450	723,450
Hospital-Month-Year FE	✓	✓	✓

Notes: Table displays results for the regression detailed in Equation (5). The dependent variable is 1 if the patient's diagnosis of delivery method deviates from the clinical consensus, measured as the patient's predicted C-section risk, 0 otherwise. We use a threshold of 20%, so that patients with a risk of 20% or below have a deviation if the physician performs a C-section, and patients with a risk of 80% or above have a deviation if the physician performs a vaginal delivery. Deliveries are restricted to claims in Medicaid 2015-2019, with active physicians (at least 100 deliveries over the sample period), for whom residency location is identified. Residencies are restricted to those with at least 10 practicing alumni and observable patient mix estimates from the ACS. * indicates $p < 0.05$, ** for $p < 0.01$, and *** for $p < .001$.

5. Discussion

In this paper, we explore the role of training in explaining racial gaps in maternal health outcomes in the US. We focus on residency in obstetrics and gynecology, using granular Medicaid claims data from across the US and exploiting both physician mobility post-residency and quasi-random allocation of patients to physicians during emergent deliveries.

Our analysis suggests that variation across residencies matters relatively little in explaining treatment and outcomes differences across patients. This result is perhaps surprising in the context of the substantial geographic variation in treatment use (including C-sections) and patient mix, both of which could shape very different residency environments. This stands in contrast to the prevailing view among practitioners we interviewed, many of whom pointed to residency programs reputed for especially high or low C-section use.

When examining the effect of training with a more diverse patient mix specifically, we found no discernible impact on the quality of diagnostics for Black patients. Again, this runs counter to efforts in medical education to advertise patient mix as an advantage of certain residencies. We also found that a residency's relative effect on C-section use or delivery complications for patients of different races was only weakly correlated; in other words, C-section intensity for White patients was a relatively poor predictor of C-section intensity for Black patients. This could reflect disparities in the treatment of patients by race, which dominate the general clinical effect from learning at residency.

Our analysis focuses on existing differences in practice style and patient mix across residencies, rather than on variation in curricula. Factors such as C-section intensity or local patient demographics likely shape the training environment, but they may not reflect explicit instruction about racial disparities in care. We observe physicians' contact with diverse patients but not whether residencies actively address race in clinical discussions. Programs that do so, by teaching how conditions vary across populations or by contextualizing health disparities, may produce more equitable outcomes. Evaluating such interventions in residency and continuing education remains an important area for future research.

Our analysis has several limitations, and also leaves open questions. Most notably, data on physician education—both who attended which residency, and residency characteristics—is extremely rare. Though we use best-in-class data and supplement that with residency information based on several administrative datasets, it remains challenging to generate and test measures of the residency environment. More systematic data collection on residency characteristics—such as cohort size and diversity, or text-based measures of residency sentiment and ethos—could shed additional light on variation that such characteristics might generate. In this vein, though we are unable to find a significant impact of training with a diverse patient mix on racial gaps in maternal health, our treatment variable for this analysis measures only contact. It may be that contact must be *mediated* to be effective; for example, through open discussion on racial disparities in healthcare and the different risks faced by patients of different races. Other forms of contact—such as diversity among peer residents—may be (more) effective.

References

AAMC. (2023, January). *2022 Physician Specialty Data Report* (tech. rep.). Association of American Medical Colleges.

Alrababah, A., Marble, W., Mousa, S., & Siegel, A. A. (2021). Can Exposure to Celebrities Reduce Prejudice? The Effect of Mohamed Salah on Islamophobic Behaviors and Attitudes. *American Political Science Review*, *115*(4), 1111–1128.

Alsan, M., Durvasula, M., Gupta, H., Schwartzstein, J., & Williams, H. (2024). Representation and Extrapolation: Evidence from Clinical Trials. *The Quarterly Journal of Economics*, *139*(1), 575–635.

Alsan, M., Garrick, O., & Graziani, G. (2019). Does Diversity Matter for Health? Experimental Evidence from Oakland. *American Economic Review*, *109*(12), 4071–4111.

Alsan, M., & Wanamaker, M. (2018). Tuskegee and the Health of Black Men*. *The Quarterly Journal of Economics*, *133*(1), 407–455.

Asch, D. A. (2009). Evaluating Obstetrical Residency Programs Using Patient Outcomes. *JAMA*, *302*(12), 1277.

Auty, S. G., Daw, J. R., Admon, L. K., & Gordon, S. H. (2024). Comparing approaches to identify live births using the Transformed Medicaid Statistical Information System. *Health Services Research*, *59*(1).

Badinski, I., Finkelstein, A., Gentzkow, M., & Hull, P. (2024). Geographic Variation in Healthcare Utilization: The Role of Physicians. *National Bureau of Economic Research*, *(No. w31749)*.

Bursztyn, L., Chaney, T., Hassan, T. A., & Rao, A. (2024). The Immigrant Next Door. *American Economic Review*, *114*(2), 348–384.

Corredor-Waldron, A., Currie, J., & Schnell, M. (2024). Drivers of Racial Differences in C-Sections [Genre: Working paper]. *National Bureau of Economic Research*, *(wp 32891)*.

Currie, J., & MacLeod, W. B. (2017). Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians. *Journal of Labor Economics*, *35*(1), 1–43.

Doyle, J. J., Graves, J. A., Gruber, J., & Kleiner, S. A. (2015). Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns. *Journal of Political Economy*, *123*(1), 170–214.

Dranove, D., Ramanarayanan, S., & Sfekas, A. (2011). Does the Market Punish Aggressive Experts? Evidence from Cesarean Sections. *The B.E. Journal of Economic Analysis & Policy*, *11*(2).

Eichmeyer, S., & Zhang, J. (2022). Pathways into Opioid Dependence: Evidence from Practice Variation in Emergency Departments. *American Economic Journal: Applied Economics*, *14*(4), 271–300.

Einiö, E., Feng, J., & Jaravel, X. (2022). Social Push and the Direction of Innovation.

Epstein, A. J., Ketcham, J. D., & Nicholson, S. (2010). Specialization and matching in professional services firms. *The RAND Journal of Economics*, *41*(4), 811–834.

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.

Fingar, K. R., Mabry-Hernandez, I., Ngo-Metzger, Q., Wolff, T., Steiner, C. A., & Elixhauser, A. (2017). *Delivery Hospitalizations Involving Preeclampsia and Eclampsia, 2005–2014* (tech. rep.). Agency for Healthcare Research and Quality.

Gørtz, M., Kristiansen, I., & Wang, T. (2024, November). *The Power of Daughters: How Physicians' Family Influences Female Patients' Health* (tech. rep. No. w33146). National Bureau of Economic Research. Cambridge, MA.

Greenwood, B. N., Carnahan, S., & Huang, L. (2018). Patient–physician gender concordance and increased mortality among female heart attack patients [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *115*(34), 8569–8574.

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.

Gunja, M., Gumas, E., Masitha, R., & Zephyrin, L. (2024, June). Insights into the U.S. Maternal Mortality Crisis: An International Comparison.

Gunja, M., Gumas, E., & Williams, R. (2023). Global Perspective on U.S. Health Care.

Hamm, R. F., Srinivas, S. K., & Levine, L. D. (2020). A standardized labor induction protocol: Impact on racial disparities in obstetrical outcomes. *American Journal of Obstetrics & Gynecology MFM*, *2*(3), 100148.

Hoyert, D. (2022, February). *Maternal Mortality Rates in the United States, 2020* (tech. rep.). National Center for Health Statistics (U.S.)

Johnson, E. M., & Rehavi, M. M. (2016). Physicians Treating Physicians: Information and Incentives in Childbirth. *American Economic Journal: Economic Policy*, *8*(1), 115–141.

Kennedy-Moulton, K., Miller, S., Wherry, L. R., Persson, P., Rossin-Slater, M., & Aldana, G. (2022). Maternal and Infant Health Inequality: New Evidence from Linked Administrative Data. *NBER Working Paper*.

Khambalia, A. Z., Roberts, C. L., Nguyen, M., Algert, C. S., Nicholl, M. C., & Morris, J. (2013). Predicting date of birth and examining the best time to date a pregnancy. *International Journal of Gynecology & Obstetrics*, *123*(2), 105–109.

Kling, J. R. (2006). Incarceration Length, Employment, and Earnings. *American Economic Review*, *96*(3), 863–876.

Main, E. K., Chang, S.-C., Tucker, C. M., Sakowski, C., Leonard, S. A., & Rosenstein, M. G. (2023). Hospital-level variation in racial disparities in low-risk nulliparous cesarean delivery rates. *American Journal of Obstetrics & Gynecology MFM*, *5*(12), 101145.

Massie, J. P., Cho, D. Y., Kneib, C. J., Sousa, J. D., Morrison, S. D., & Friedrich, J. B. (2021). A Picture of Modern Medicine: Race and Visual Representation in Medical Literature. *Journal of the National Medical Association*, *113*(1), 88–94.

Philip, M., & Ozkaya, O. (2024). Disparate Treatment and Outcomes in Emergency Departments: Evidence from Florida.

Phillips, R. L., Petterson, S. M., Bazemore, A. W., Wingrove, P., & Puffer, J. C. (2017). The Effects of Training Institution Practice Costs, Quality, and Other Characteristics on Future Practice. *The Annals of Family Medicine*, *15*(2), 140–148.

Robinson, S., Royer, H., & Silver, D. (2023, November). Geographic Variation in Cesarean Sections in the United States: Trends, Correlates, and Other Interesting Facts.

Schnell, M., & Currie, J. (2018). Addressing the Opioid Epidemic: Is There a Role for Physician Education? [Publisher: The University of Chicago Press]. *American Journal of Health Economics*, *4*(3), 383–410.

Stanford Center for Population Health Sciences. (2023a). Medicaid 100% [2011-2019] Other Services (OT) (Version 3.0) [Dataset] [Redivis (DOI:10.71778/V2DW-7A53). Available at https://doi.org/10.57761/8hk4-0994].

Stanford Center for Population Health Sciences. (2023b). Medicaid 100% [2011–2019] Inpatient (IP) (Version 2.0) [Dataset] [Redivis (DOI:10.71778/V2DW-7A53). Available at https://doi.org/10.57761/9mte-t021].

Turner, C., Reed, K., & May, R. (2010, June). *The Seventh Annual HealthGrades Maternity Care in American Hospitals Study* (tech. rep.). Health Grades, Inc.

US Preventive Services Task Force, Davidson, K. W., Barry, M. J., Mangione, C. M., Cabana, M., Caughey, A. B., Davis, E. M., Donahue, K. E., Doubeni, C. A., Kubik, M., Li, L., Ogedegbe, G., Pbert, L., Silverstein, M., Simon, M. A., Stevermer, J., Tseng, C.-W., & Wong, J. B. (2021). Aspirin Use to Prevent Preeclampsia and Related Morbidity and Mortality: US Preventive Services Task Force Recommendation Statement. *JAMA*, *326*(12), 1186.

A. Measurement

A.1 Delivery Complications

To measure delivery complications, we use two sources. We begin with a list of complications used by Asch, 2009 to identify delivery complications based on clinical advice, the ICD-9 diagnosis codes for which are presented in Table A1. We supplement this with a list of delivery complications from Health Grades, Inc., used to measure hospital quality in maternal care (Turner et al., 2010). This measure has previously been used in Epstein et al., 2010. We summarize the additional diagnosis codes unique to Turner et al., 2010 in Table A2.

In constructing our outcome indicator, our goal is to measure a reliable signal of physician performance. We therefore want to be comprehensive regarding possible complications, but specific to those a physician could plausibly influence. While the set of diagnosis codes in Turner et al., 2010 is more complete, it also covers more elements of the delivery process that providers other than the OB/GYN attending the delivery are responsible for. For example, it includes anesthetic complications, but anesthetic procedures will typically be provided by an anesthesiologist. Similarly, catheters and transfusions might be provided by nurses or other assistants during labor.

After combining and categorizing the diagnosis codes used in each measure, we therefore make a couple of adjustments. First, we exclude diagnosis codes related to complications from catheters, transfusions, injections, or anesthesia that are present in Turner et al., 2010 but not Asch, 2009. Second, we refine the inclusion of codes for lacerations or perineal tears. We exclude any codes that do not specify the location (e.g., cervical, high vaginal) or degree, and we exclude first- and second-degree perineal tears. The latter are exceedingly common (as many as 90% of nulliparous patients undergoing vaginal delivery will experience one) and therefore a less reliable signal of physician performance.

We translate the ICD-9 codes to ICD-10 codes manually, and search the diagnosis codes associated with each delivery claim in both Medicaid's Inpatient (IP) and Other Services (OT) files. The final codes used in the construction of our delivery complication measure are available in Table A3.

Table A1: Complication Diagnosis Codes by Delivery Mode from Asch, 2009

Complication	**Diagnosis Codes (ICD-9)**
Vaginal	
Hemorrhage	287.4, 648.22, 649.32, 666.02, 666.04, 666.10, 666.12, 666.14, 666.20, 666.22, 666.24, 666.32, 667.02, 667.12, 998.11
Laceration	664.21, 664.31, 664.41, 664.51, 664.61, 664.81, 664.91, 665.31, 665.41, 674.22
Infection	646.62, 670.02, 672.02, 996.60, 996.62, 999.3
Other	512.0, 512.1, 512.8, 518.4, 518.81, 584.5, 584.8, 584.9, 665.51, 669.12, 669.22, 669.42, 669.82, 671.42, 671.52, 673.12, 673.22, 674.32, 674.52, 674.92, 785.50, 785.51, 785.59, 996.31, 998.4, 998.7, 999.2, 999.4, 999.5, 999.6, 999.7, 999.8, 999.9
C-Section	
Hemorrhage	287.4, 648.22, 649.32, 666.02, 666.04, 666.10, 666.12, 666.14, 666.20, 666.22, 666.24, 666.32, 667.02, 667.12, 998.11
Infection	646.62, 670.02, 672.02, 674.12, 674.32, 998.59, 999.3
Operative	669.42, 998.20, 998.4, 999.83
Other	512.0, 512.1, 512.8, 518.4, 518.81, 584.5, 584.8, 584.9, 669.12, 669.22, 669.82, 671.42, 671.52, 673.12, 673.22, 674.52, 674.92, 785.50, 785.51, 785.59, 996.31, 996.60, 996.62, 998.60, 998.7, 999.2, 999.4, 999.5, 999.6, 999.7, 999.8, 999.9

Notes: Table details ICD-9 codes used by Asch, 2009 to identify delivery complications for vaginal and C-section delivery.

Table A2: Additional Complication Diagnosis Codes from Turner et al., 2010

Complication	**Diagnosis Codes (ICD-9)**	
All Delivery Types		
	666.30, 666.34	Postpart. coagulation defect
	668.02, 668.12, 668.14, 668.22, 668.82	Complex. of anesthetic
	997.0, 997.0x	Nervous system complex.
	997.1	Cardiac complex.
	997.3	Respiratory complex.
	997.4	Digestive system complex.
	997.5	Urinary complex.
	997.91	Complications unclassified
	998.0	Postop. shock
	998.3x	Disruption of wound
	998.81	Emphysema
	998.83	Non-healing wound
	998.9	Unspecified complex.
	999.1	Air embolism
	999.31	Infection from catheter
	999.39	Infection from infusion etc.
Vaginal		
	665.1, 665.11	Rupture of uterus
	665.22	Inversion of uterus
	669.14	Shock
	998.2	Accidental laceration
	998.59	Other post-op. infection
C-Section		
	666.00	Postpart. hemorrhage
	668.04, 668.24	Complex. of anesthetic
	669.44	Other complex. of obstetric surgery
	998.51	Infected postop. seroma
	998.7	Reaction to foreign substance left in procedure

Notes: Table details ICD-9 codes used exclusively by Turner et al., 2010, and not Asch, 2009. We abbreviate “complications” to “complex.” An “x” within ICD codes indicates any value can be taken.

Table A3: All Complication Diagnosis Codes

Complication	Diagnosis Codes (ICD-9 and ICD-10)
Vaginal	
Hemorrhage	**ICD-9:** 287.4, 648.22, 649.32, 666.00, 666.02, 666.04, 666.10, 666.12, 666.14, 666.20, 666.22, 666.24, 666.32, 667.02, 667.12, 998.11 **ICD-10:** D69.6, O46.9, O99.19, O72.x
Laceration	**ICD-9:** 664.21, 664.31, 664.51, 664.61, 665.31, 665.41 **ICD-10:** O70.2, O70.3, O71.3, O71.4, O71.5, O71.89, O71.9
Infection	**ICD-9:** 646.62, 670.02, 672.02, 996.60, 996.62, 999.3 **ICD-10:** O23.4, O85, O86.8, T85.79XA, T80.2XXA, O23.5, O86.89
Other	**ICD-9:** 512.0, 512.1, 512.8, 518.4, 518.81, 584.5, 584.8, 584.9, 665.10, 665.11, 665.22, 665.51, 666.30, 666.34, 669.12, 669.14, 669.22, 669.42, 669.82, 671.42, 671.52, 673.12, 673.22, 674.32, 674.52, 674.92, 785.50, 785.51, 785.59, 996.31, 997.00, 997.1, 997.3, 997.4, 997.5, 997.91, 998.0, 998.2, 998.3x, 998.4, 998.59, 998.7, 998.81, 998.83, 998.9, 999.2, 999.4, 999.5, 999.6, 999.7, 999.8, 999.9 **ICD-10:** D78.12, E36.12, G97.49, G97.81, I97.3, I97.52, I97.88, J93.0, J95.72, J95.811, J93.83, J81.0, J96.00, L76.12, K68.11, K91.72, N17.1, N17.8, N17.9, N99.72, N99.89, O71.1, O71.2, O75.1, O75.89, O62.9, O88.11, O88.22, O22.2, O22.3, O88.01, O88.22, O90.3, O90.89, O90.9, R57.9, R65.21, R57.8, T80.5XXA, T80.6XXA, T80.89XA, T80.89XA, T80.9XXA, T81.30XA, T81.31XA, T81.32XA, T81.4XXA, T81.49XA, T81.515A, T81.82XA, T81.89XA, T81.9XXA, T82.01XA, T88.7XXA, T88.8XXA
C-Section	
Hemorrhage	**ICD-9:** 287.4, 648.22, 649.32, 666.00, 666.02, 666.04, 666.10, 666.12, 666.14, 666.20, 666.22, 666.24, 666.32, 667.02, 667.12, 998.11 **ICD-10:** D69.6, O46.9, O99.19, O72.x
Infection	**ICD-9:** 646.62, 670.02, 672.02, 674.12, 674.32, 998.51, 998.59, 999.3 **ICD-10:** O23.4, O23.5, O85, O86.89, O90.1, O90.3, T80.2XXA, T81.41XA, T81.4XXA
Operative	**ICD-9:** 669.44, 669.42, 998.3x, 998.20, 998.4, 998.7, 998.83, 999.83 **ICD-10:** O75.4, O88.11, T81.30XA, T81.31XA, T81.32XA, T81.2XXA, T81.515A, T81.89XA, T80.0XXA
Other	**ICD-9:** 512.0, 512.1, 512.8, 518.4, 518.81, 584.5, 584.8, 584.9, 666.30, 666.34, 669.12, 669.22, 669.82, 671.42, 671.52, 673.12, 673.22, 674.52, 674.92, 785.50, 785.51, 785.59, 996.31, 996.60, 996.62, 997.00, 997.1, 997.3, 997.4, 997.5, 997.91, 998.0, 998.60, 998.7, 998.81, 998.9, 999.2, 999.4, 999.5, 999.6, 999.7, 999.8, 999.9 **ICD-10:** G97.81, I97.3, I97.88, J93.0, J95.811, J93.83, J81.0, J96.00, N17.1, N17.8, N17.9, N99.89, O75.1, O62.9, O75.89, O22.22, O22.3, O88.01, O88.22, O90.89, O90.9, R57.9, R65.21, R57.8, T81.61XA, T81.69XA, T81.82XA, T81.9XXA, T82.01XA, T85.79XA, T81.49XA, T80.5XXA, T88.7XXA, T80.89XA, T80.9XXA

B. Additional Tables and Figures

Table A4: Covariates Predicting C-Section Appropriateness

	This Paper	Currie and MacLeod, 2017	Robinson et al., 2023
Maternal Age	5-year bins	5-year bins	5-year bins
Term (37+ weeks)	✓	–	✓
Prenatal Visits	≥ 19	–	≥ 19
Nulliparous	✓	Birth order	✓
Singleton	✓	Multiples	✓
Vertex	✓	Breech	✓
Growth Restrictions	✓(intrauterine)	–	✓
Eclampsia	✓	–	✓
Preeclampsia	✓	–	✓
Other Hypertension	✓	✓	–
Asthma	✓	Chronic lung condition	–
Diabetes	✓	–	✓
Obesity	✓	–	–
Placenta Previa	✓	✓	–
Placental Abruption	✓	✓	–
Herpes	✓	✓	–
Hydramnios	✓	✓	–
Chorioamnionitis	✓	–	–
Cord Prolapse	✓	✓	–
Isoimmunization	✓	✓(Rh Sensitivity)	–
Macrosomia	✓	–	–
Antepartum Hemorrhage	✓	Uterine Bleeding	–
Previous C-Section	✓	✓	–
Previous Large Infant	–	✓	–
Previous Preterm	–	✓	–
Cardiac condition	✓+ congenital	✓	–
Blood disorder	✓	Anemia + Hemoglobinopathy	–
Cervical incompetence	✓	✓	–
Renal abnormalities	✓	✓	–

Notes: Table compares covariates used for prediction of delivery method in this paper, Currie and MacLeod, 2017, and Robinson et al., 2023. Age is classified into 5-year bins as follows: $<$20, 20-24, 25-29, 30-34, $\geq$35. Common covariates marked with a ✓. Covariates measuring the same variable with different definitions (e.g., nulliparous versus birth order) are noted in text.

Table A5: Logistic Regression Model of C-Section Risk

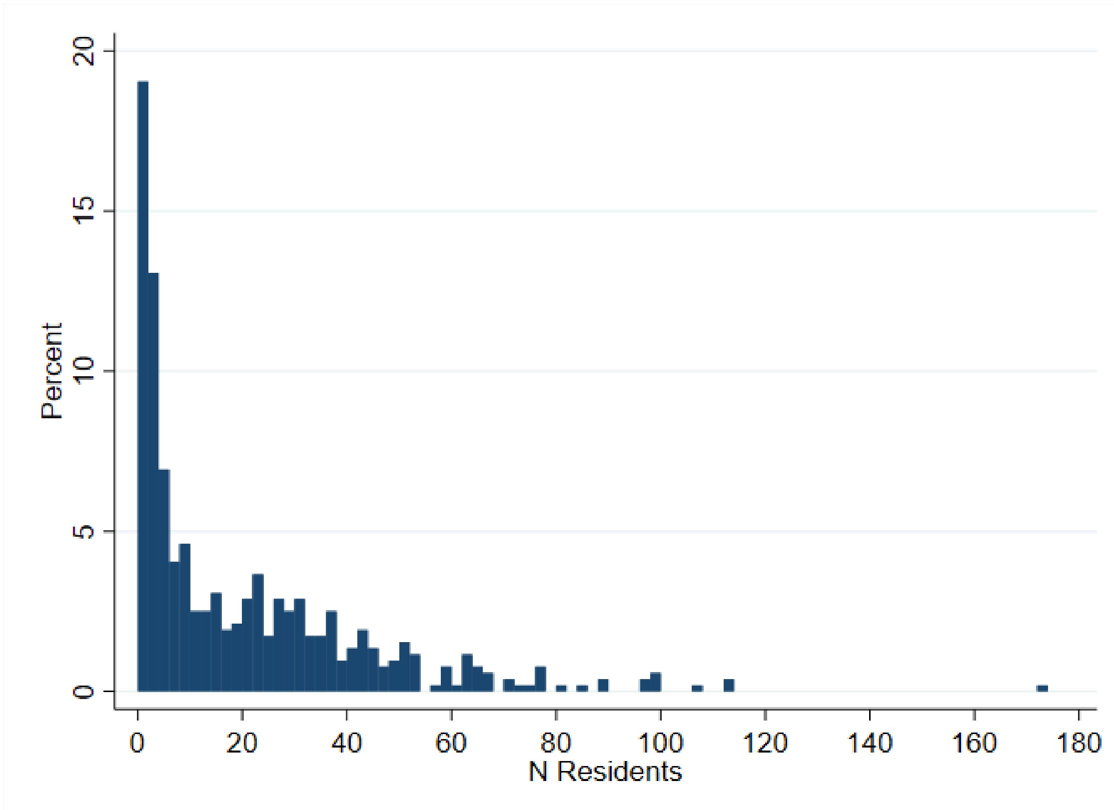
	Beta	SE
Age < 20	-.1928386***	.004246
Age 25-29	.0288876***	.0027257
Age 30-34	.0893298***	.0030957
Age 35+	.2269394***	.0035983
Prenatal Claims ≥ 19	.1046772***	.0021831
Nulliparous	.7624481***	.0026043
Term	.0072586	.0048402
Singleton	-.6099307***	.0040694
Vertex	-3.056423***	.0057432
Intrauterine Growth Restriction	.5275533***	.005088
Eclampsia	1.432985***	.0264771
Preeclampsia	.8974775***	.0040323
Other Hypertension	.3488085***	.0036775
Asthma	-.0124584**	.0050635
Diabetes	.3131867***	.003878
Obesity	.4381966***	.0033706
Placenta Previa	1.426388***	.0112711
Placental Separation	1.443811***	.0083788
Herpes	.2789016***	.0058824
Hydramnios	.4656481***	.0048721
Chorioamnionitis	.8671536***	.0058849
Cord Complications	-.481389***	.0029228
Isoimmunization	-.0694652***	.0105647
Macrosomia	1.696111***	.0062682
Antepartum Hemorrhage	.7995364***	.0166766
Previous C-Section	3.744066***	.002855
Congenital Heart Disease	.4093788***	.0125014
Blood Disorder	.1283061***	.0028281
Cervical Incompetence	.1501317***	.0134766
Renal Abnorm.	.3903928***	.0242845
Constant	1.310074***	.0079258
Observations	7820852	

Notes: Table describes the coefficients and standard errors from a logistic regression of C-section delivery (a 0/1 dependent variable) on patient covariates. * indicates $p < 0.05$, ** for $p < 0.01$, and *** for $p < .001$. The sample includes all deliveries identified in Medicaid claims, 2015-2019.

Sample Descriptives on Residents and Births in Medicaid by Residency

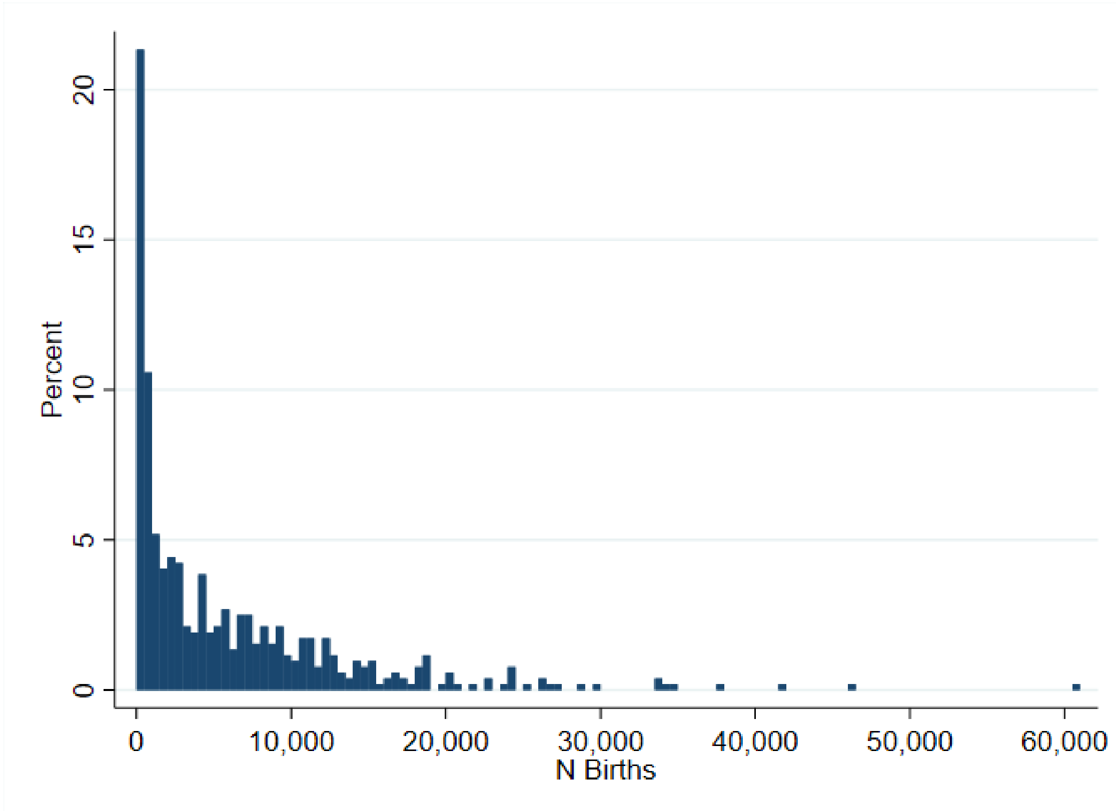
In Figure A1 we plot the distribution of the number of alumni active in Medicaid claims, 2015-2019, for each of the 520 residencies linked from the AMA MasterFile. The mean is 20, though the distribution is skewed, with 25% of residencies having 2 or fewer residents practicing in Medicaid. In Figure A2 we plot the distribution of the number of births per residency identified in Medicaid claims, 2015-2019. The mean number of births per residency over this period is nearly 6,000. The number of alumni and number of births identified in the Medicaid claims is directly proportional (with a correlation coefficient of $\rho = 0.96$).

Figure A1: Distribution of Number of Alumni per Residency in Medicaid Claims, 2015-2019



Notes: Histogram shows the distribution of the number of residency alumni (physicians) per residency actively practicing in obstetrics in Medicaid claims, 2015-2019. The 25^{th} percentile is 2, mean is 19.99 residents, standard deviation 22.96, minimum 1, and maximum 173. $N = 520$ residencies.

Figure A2: Distribution of Number of Births per Residency in Medicaid Claims, 2015-2019



Notes: Histogram shows the distribution of the number of births per residency, among physicians actively practicing in obstetrics in Medicaid claims, 2015-2019. The mean number of births per residency is 5,959, standard deviation 7,514, minimum 102, and maximum 60,787. $N = 520$ residencies.

An Additional Measure of Physician Deviation in Delivery Method Choice

In Section 4.2, we assess the impact of training with a more diverse patient mix on the differential rate of deviations from medical consensus in delivery method. Here, we present an alternative measure of those deviations. We use the same specification, as follows:

$$\begin{aligned} \text{Sq. Deviation}_{ijbt} = & \alpha_0 + \beta_0 \text{Black}_i + \beta_1 \text{ResidencyPatientMix}_j \\ & + \beta_2 (\text{Black}_i \times \text{ResidencyPatientMix}_j) \\ & + \gamma_{bt} + \varepsilon_{ijbt} \end{aligned} \tag{6}$$

Where Sq. Deviation_{it} is the squared error in patient i's delivery method performed by physician j at hospital b in month-year t, calculated as the indicator for C-section minus the predicted C-section risk, all squared. Fixed effects control for the assignment of patient to physician, as well as seasonality in births. Standard errors are clustered at the physician level. As above, we focus on the Black-White race gap, and so restrict to patients who identify as either Black or White.

Black_i is an indicator equal to one if the patient is Black, zero otherwise. The coefficient β_0 captures the differential squared deviation in Black patients' delivery method diagnoses (relative to non-Hispanic White patients). We analyze how that differential squared deviation changes when we include patient mix at physician residency as a covariate (ResidencyPatientMix$_j$), and the interaction of the two.

Our results, shown in Table A6, show that Black patients are more likely to be misdiagnosed in delivery method. But the training experience of their physician with a more diverse patient mix has little discernible impact on that diagnostic error.

Table A6: The Impact of Residency Patient Mix on Diagnosis Accuracy by Patient Race

	Individual SE	Individual SE	Individual SE
Black	0.0092***	0.0092***	0.0110***
	(0.0007)	(0.0007)	(0.0012)
Residency		0.0031	0.0059**
Patient Mix		(0.0026)	(0.0029)
Black ×			-0.0073*
Residency Patient Mix			(0.0038)
Constant	0.1161***	0.1154***	0.1147***
	(0.0003)	(0.0007)	(0.0008)
Mean (All)	0.1195	0.1195	0.1195
Mean (White)	0.1138	0.1138	0.1138
# NPIs (Clusters)	6,632	6,632	6,632
N	723,450	723,450	723,450
Month-Year FE	✓	✓	✓
Hospital FE	✓	✓	✓
Physician FE	✓	✓	✓

Notes: Table displays results for the regression detailed in Equation (6). The dependent variable is the error in a patient's diagnosis of delivery method, calculated as an indicator for C-section minus the patient's C-section risk, all squared. Deliveries are restricted to claims in Medicaid 2015-2019, with active physicians (at least 100 deliveries over the sample period), for whom residency location is identified. Residencies are restricted to those with at least 10 practicing alumni and observable patient mix estimates from the ACS. * indicates $p < 0.05$, ** for $p < 0.01$, and *** for $p < .001$.