Geographic Variation in Cesarean Sections in the United States: Trends, Correlates, and Other Interesting Facts *

Sarah Robinson
Claremont McKenna College
sarah.robinson@claremontmckenna.edu

Heather Royer
UC Santa Barbara & IZA & NBER
heather.royer@ucsb.edu

David Silver
UC Santa Barbara & NBER
dsilver@ucsb.edu

This version: October 30, 2023

Abstract

Analyzing data spanning three decades covering the near universe of births, we study county-level differences in Cesarean section (C-section) rates among first-time mothers of singleton births. Our research reveals persistent geographic variation in C-section rates for both low- and high-risk groups. Counties with elevated C-section rates consistently perform more C-sections across mothers at all levels of appropriateness for the procedure. These elevated rates of C-section in high C-section counties are associated with reduced maternal and infant morbidity. We also find that C-section decisions are less responsive to underlying risks for Black mothers relative to white mothers, suggesting potential welfare-reducing disparities.

*We thank David Card for inspiring us to write this paper. We thank Juliana Dajon and Nicolas Fuertes Segura for assisting in the research. We thank Janet Currie, Stefanie Fischer, Thomas Lemieux, Marit Rehavi, Corey White, participants at DaveFest, and UCSB seminar participants for their constructive comments and feedback. Sarah Robinson thanks NBER's National Institute on Aging, Grant Number T32-AG000186 for financial support.

1 Introduction

Cesarean section (C-section) is the most common surgical procedure performed in the United States. The decision to proceed with a C-section involves considerable provider discretion, as the conditions precipitating the procedure's use are multi-faceted, and because the consequences of unnecessary C-sections are highly-debated (e.g., Hyde et al., 2012; Shah, 2017; Keag, Norman, and Stock, 2018).[1] The complexity of this decision process poses challenges for healthcare providers in determining the appropriate use of C-sections. As a consequence, providers sometimes resort to simple heuristics to guide their decisions, often resulting in worse outcomes (e.g., Singh, 2021). Due to the inherent ambiguity in determining the appropriateness of a C-section, it is not surprising that C-section rates vary widely across geographies. For instance, in 2017, the 95$^{\text{th}}$ percentile of county-level C-section rates was 39.4 percent, whereas the 5$^{\text{th}}$ percentile recorded a rate of 24.4 percent.[2]

These patterns are well-situated within a literature documenting geographic variation in health care inputs in numerous other domains, from tonsillectomies in children (Glover, 1938), to catheterization procedures in heart attack patients (e.g., Molitor, 2018), and to medical spending more generally (e.g., Wennberg and Gittelsohn, 1973; Finkelstein, Gentzkow, and Williams, 2016). Often, regions with higher utilization produce outcomes equal or inferior to their lower-utilization counterparts, a pattern dubbed "flat-of-the-curve medicine" (e.g., Fuchs, 2004). This phenomenon naturally raises the question: Is the utilization of intensive healthcare inputs inefficient in areas with high usage, such that reducing interventions could improve welfare? In the context of childbirth, this line of reasoning has fueled a push to reduce the perceived overuse of

[1] Conditions precipitating the use of C-section include stalled labor, distressed infant, breech birth, multiple birth pregnancy, placenta previa, prolapsed umbilical cord, and a previous pregnancy delivery via C-section. Placenta previa is a condition whereby the placenta covers the cervix, and a prolapsed umbilical cord occurs when the umbilical cord precedes the infant's head during delivery. See https://www.mayoclinic.org/tests-procedures/c-section/about/pac-20393655.

[2] Based on authors' calculations.

C-sections. For example, the Centers for Disease Control & Prevention's public health objectives, outlined in *Healthy People 2030*, include a target of reducing C-section rates for low-risk first births to 23.6 percent, while the actual rate in 2018 stood at 25.6 percent.[3] Relatedly, in California, the California Health Care Foundation, in cooperation with the Robert Wood Johnson Foundation, funded an initiative aimed at reducing unnecessary C-section through the provision of transparent hospital performance metrics and decision aids.[4]

In this study, we build upon previous research documenting geographic variation in C-section rates (e.g., Baicker, Buckles, and Chandra, 2006; Epstein and Nicholson, 2009; Kozhimannil, Law, and Virnig, 2013). We examine geographic variation and trends in the use of C-section over nearly three decades, from 1989 to 2017, during which period the national C-section rate increased from 23 percent to 32 percent (author's calculations). We characterize this variation using insights from Currie and MacLeod (2017) and Chandra and Staiger (2007) who provide frameworks for evaluating the appropriateness of health care utilization. Our goal is to understand the role of place in the use of C-sections. We aim to contribute further evidence to the ongoing debate about whether C-section rates could be reduced without adverse welfare consequences.

The domain of maternal and infant health provides a rich environment for the study of geographic variation in health care and health outcomes, in part due to the availability of data. In particular, birth and infant death data published annually by the National Vital Statistics System covers the near universe of births occurring in the United States. These data are advantageous for a number of reasons, including large sample sizes, and full geographic and demographic coverage.[5] Our analytic sample is first-time mothers

[3] Actual rate based on authors' calculations.

[4] See `https://www.chcf.org/project/reducing-unnecessary-c-sections/`.

[5] Much of the existing literature on geographic variation in healthcare has concentrated on the Medicare population, and in particular, those enrolled in traditional fee-for-service Medicare. Medicare eligibles have increasingly opted for privately administered Medicare Advantage plans, for which utilization data are not centrally collected. Thus, studies of recent data must rely on an increasingly selected sample of the Medicare population; around half of Medicare-eligible individuals now enroll in Medicare Advantage (Trish et al., 2023). This selection issue notwithstanding, the Medicare setting – consisting of univer-

with a singleton birth. Focusing on this sample avoids the selection problems inherent in second- and higher-order births due to the strong impact of a previous C-section delivery on delivery choice in subsequent births.

Our analysis is divided into two parts. In the first part, we document patterns in county-level C-section rates. We focus on three relationships: temporal trends, geographic variations, and county-level associations between risk-adjusted C-section rates and outcomes for both infants and mothers. Much of the effort to reduce C-section rates focuses on births with low risk of poor birth outcomes, for which C-sections are considered less necessary and hence more prone to overuse. In light of this suspected overuse, our analysis dissects these relationships separately for low- and high-risk births.[6]

In the second part of the paper, we peer inside these county-level aggregates to detail the relationship between C-section *appropriateness*[7] and C-section delivery across counties categorized by their overall C-section rates. We also relate this measure of C-section appropriateness to health outcomes, allowing us to provide insights into whether higher rates of C-section lead to worse outcomes, distinguishing between mothers for whom C-sections are most appropriate and those for whom they are less suitable (the population of focus in much of the public discourse and in policy interventions). Given the documented disparities in care across racial groups, we also investigate racial differences in these associations.

Our analysis yields several key findings. First, while C-section rates rose substantially from the late 1990s through 2010, rates have plateaued over the past decade. This plateauing reflects divergent trends among high- and low-risk births in our sample of

sal coverage of care for the elderly and disabled – may be limited in its ability to speak to other domains of interest. See, for instance, (Cooper et al., 2019).

[6]We follow policymakers who have focused on the set of nulliparous term singleton vertex (NTSV) births as "low risk." As stated earlier, our analytic sample limits to nulliparous (first-time) mothers with singleton (non-twin, triplet, etc.) births. Within this sample, we classify births as low risk if they are also at term and vertex ("term" means gestation of at least 37 weeks, and "vertex" means non-breech). Additionally, to be conservative, we consider four medical risk factors (eclampsia, pre-eclampsia, diabetes, and growth restrictions), as well as maternal age and prenatal visits, in our classification of low-risk births, as laid out in Table 1.

[7]We estimate appropriateness based on a model including medical and demographic covariates.

singleton first births. C-sections among high-risk mothers continued to increase in the 2010s, albeit at a considerably slower pace compared to the surge experienced in the 2000s. Conversely, among low-risk singleton first births, C-section rates started to fall beginning in 2010.

Second, geographic variation in C-section is persistent. A county with a 10 percentage point higher C-section rate at the outset of our study period (1989-1991) maintained, on average, a 4.4 percentage point higher C-section rate for high-risk births and a 4.7 percentage point higher rate for low-risk births at the conclusion of our study period (2015-2017). Third, counties performing more C-sections for high-risk births also tend to perform more C-sections for low-risk births —a 10 percentage point increase in high-risk C-section rates is associated with a 7-8 percentage point increase in low-risk C-section rates. Fourth, counties with higher C-section rates demonstrate significantly lower rates of infant and maternal morbidity, with no statistically significant difference in infant mortality. These relationships are particularly surprising in light of the fact that the populations in higher C-section counties tend to be less socioeconomically advantaged (e.g., lower income and education, higher unemployment and Medicaid coverage) and would thus be expected to have higher morbidity rates, all else equal.

Fifth, turning to our appropriateness measures, high C-section and low C-section counties respond similarly to measures of C-section appropriateness, although high C-section counties have higher rates of C-section across the appropriateness distribution. Despite the greater propensity for surgical intervention, the incidence of poor outcomes in high C-section counties is by many measures lower than that in low C-section counties across the appropriateness distribution (not just on average, as pointed out above). Finally, turning to an analysis of race, we find that non-Hispanic Black mothers with the highest measured C-section appropriateness receive C-sections *less* frequently than similarly-appropriate non-Hispanic white mothers, yet have *higher* rates of neonatal mortality and infant and maternal morbidity, suggestive of misallocation of C-sections among Black mothers.

Collectively, our correlational results are consistent with geographic variations in C-section driven by place-based differences in surgical skill (i.e., skill in performing a C-section) rather than differences in diagnostic skill (i.e., skill in targeting interventions to the most appropriate patients; Currie and MacLeod (2017)). Under a model of surgical skill, reductions in C-section rates among low-risk births in high C-section places could unintentionally lead to reductions in C-sections among mothers and infants who may genuinely benefit from the procedure. Furthermore, our results by racial group highlight possible welfare-reducing disparities in surgical intervention for Black mothers, especially those we estimate to be most appropriate for C-section delivery.

One persistent challenge in estimating causal effects of places on healthcare delivery is the endogenous sorting of providers and patients across geographies. Recent advances in related research studying the Medicare setting use patients and providers who move across places to tease apart differences due to selection versus place itself (e.g., Finkelstein, Gentzkow, and Williams, 2016; Molitor, 2018). While our paper does not provide a fully satisfactory solution to this fundamental issue, as we are unable to track patients or providers across places in our data, we control for a wealth of observable characteristics in an effort to address selection.[8] Thus, all of our analysis should be interpreted as conditional correlations rather than causal effects. Nevertheless, the compendium of descriptive statistics presented in this paper broadens the understanding of C-section usage across the United States.

Our study contributes to the extensive body of literature on geographic disparities in healthcare access and health outcomes. This literature has largely concentrated on the Medicare population (e.g., Wennberg and Gittelsohn, 1973; Chandra and Staiger, 2007; Finkelstein, Gentzkow, and Williams, 2016; Molitor, 2018; Deryugina and Molitor, 2020;

[8] Mothers with multiple births in different places could theoretically help to resolve this selection issue, although any such estimates would need to account for the high path dependence in delivery mode, and would be limited to mothers experiencing two or more births. For a recent analysis of this sort examining birthweight variation across zip codes in California, see Chyn and Shenhav (2022).

Finkelstein, Gentzkow, and Williams, 2021). Medicare's national coverage and tracking of objective and relevant health metrics, namely mortality, make it well-suited for exploring the role of place.

Place-based differences in infant & maternal health are also dramatic, albeit less studied. An important exception is Baicker, Buckles, and Chandra (2006), which we discuss in more detail in the next section. Recent research on maternity ward closures (Battaglia, Forthcoming; Fischer, Royer, and White, Forthcoming) hint at a role of place-based factors in C-section usage. The setting of childbirth shares some important similarities with Medicare, offering nationwide coverage and focusing on an at-risk population. Geographic variations in childbirth are not only important in their own right, but also have implications for later-life outcomes due to the lasting effects of early-life health (Almond and Currie, 2011) and the sizable impact of improvements in infant health on overall mortality (Cutler, Deaton, and Lleras-Muney, 2006).

2 Geographic Variation in Health Care and the Appropriateness of C-Section

Our work draws on findings and insights from three related studies: Baicker, Buckles, and Chandra (2006), Currie and MacLeod (2017), and Chandra and Staiger (2007). We discuss these studies below to provide a collection of findings, frameworks, and mechanisms that will help guide the interpretation of our set of facts.

First, Baicker, Buckles, and Chandra (2006) examine geographic variation in C-section utilization among normal and low-birthweight infants across the 198 most populous U.S. counties from 1995 to 1998. They find health system characteristics such as provider density, capacity, and malpractice liability to be the strongest correlates of C-section usage, rather than patient characteristics. They conclude that higher C-section areas perform C-sections on less appropriate patients, without producing lower mortality rates.

Their results are consistent with a model of physician decision-making whereby physicians rank patients by C-section appropriateness and perform C-sections until reaching some threshold defined by outside influences (i.e., non-focal-patient characteristics), such as the availability of health care services and malpractice policies, rather than the decision being dictated by allocative efficiency. Our study complements Baicker, Buckles, and Chandra (2006) in the following ways: we examine geographic differences in the use of C-sections across the distribution of patient appropriateness; we consider more health outcomes that are less extreme than mortality; and we expand the timespan and geographic coverage of the analysis. In particular, our sample covers nearly three decades and includes a broader set of counties. By excluding less populous areas, the geographic variation in C-section rates is much smaller in Baicker, Buckles, and Chandra (2006) (a standard deviation of 3 percentage points) than in our analysis (a standard deviation of 5 percentage points).

Next, the work of Currie and MacLeod (2017) leads to further insights about C-section appropriateness. They use detailed birth certificate data from New Jersey between 1997 and 2006 to model physician treatment decisions as both a function of diagnostic and surgical skill. Diagnostic skill measures the degree to which the physician makes the correct decision for delivery, whereas surgical skill reflects how well the physician performs C-sections relative to vaginal delivery. A physician with higher diagnostic skill responds more to the underlying characteristics of the patient, essentially improving the targeting of intensive treatments. Currie and MacLeod (2017) conclude that an increase in *diagnostic* skill leads to a decline in C-sections among low-risk mothers but a rise in C-sections with better outcomes for high-risk mothers. On the other hand, an increase in *surgical* skill raises C-section rates for all. In our analysis, we see that high C-section areas tend to have elevated C-section rates across the distribution of patient appropriateness, consistent with differences in surgical skill. On the other hand, our examination of patient race suggests that diagnostic skills are also at play, as we find the treatment of Black mothers to be less sensitive to their need.

Finally, moving outside of the healthcare of childbirth, Chandra and Staiger (2007) provide an alternative explanation to the frequent "flat of the curve" observation in the healthcare literature (i.e., areas that deliver more intensive medical care have the same or worse health outcomes). They adopt a model whereby there are two treatment options —invasive and non-invasive —and the physician maximizes utility for each patient, which is a function of cost and survival. Survival from a particular procedure is a function of the fraction of individuals in the area who undergo that procedure. This particular modeling assumption leads to productivity spillovers that are harmful to patients better suited for non-invasive care. Such spillovers could occur because of physician learning, a phenomenon backed by several studies dating back at least to Coleman, Katz, and Menzel (1957).

Both Currie and MacLeod (2017) and Chandra and Staiger (2007) share the feature that a uniform decrease in the rate of intervention cannot be Pareto improving, as appropriate cases will either receive interventions less often or, when they do receive the intervention, their provider will be less skilled in performing it. These models highlight the importance of examining treatments and outcomes across the distribution of patient appropriateness in understanding variation across providers, across space, or across populations.[9]

[9] Chandra and Staiger (2020) expand on their earlier paper by studying who is treated intensively for heart attack within hospitals, focusing on the distinction between allocative inefficiency and comparative advantage. Allocative inefficiency in that framework incorporates the notion that geographic variation arises when some areas perform too much intensive care while other areas perform too little. On the other hand, comparative advantage could lead to geographic variation if some areas are better at performing more intensive care. Their main conclusion is that much of hospitals' treatment overuse is due to incorrect beliefs about the benefits of treatment. Related work in this vein includes Chan, Gentzkow, and Yu (2022), Abaluck et al. (2016), Silver (2021) and Mullainathan and Obermeyer (2022), all of which provide evidence of inefficiencies in the allocation of medical care underlying standard patterns of variation in treatment and testing rates across places and providers.

3 Data

3.1 Detailed Natality Files

Our primary data source is the Detailed Natality Files for the United States for the period 1989 to 2017, provided by the National Vital Statistics System.[10] These data cover the universe of births in the United States. They are derived from birth certificates and contain information on demographic characteristics of the parents, infant conditions (e.g., birthweight, gestational length), and procedure use (e.g., C-section, induction, ventilator, neonatal intensive care unit usage). We use the restricted version of these data with county of occurrence and county of residence information, which is the most granular geographic information available on the national data files. We tabulate our county-level statistics by county of residence to limit endogeneity concerns stemming from mothers travelling to other counties for care they prefer.[11]

We make several sample restrictions. First, since our focus is on C-section use and there is considerable path dependency in C-section use (i.e., many hospitals require that mothers deliver via C-section for higher-order births if they had a C-section for an earlier birth), we narrow our sample to first-birth mothers for our main analyses. Second, we include only singleton births in the main analysis sample, as the risk factors associated with multiple births are distinct. Third, very small counties experience very few births each year, and thus their rates of C-section are quite variable from year to year —largely due to noise. We exclude counties that ever have fewer than 100 births in a year.[12]

[10] We exclude more recent years because at the start of our project, the most recent infant mortality data was for 2017.

[11] For 75 percent of our sample, the county of occurrence and the county of residence are identical.

[12] Alternatively, one might prefer a selection criteria based on population size since we study births. If we consider all counties with a population exceeding 9,091 persons for all years (i.e., a threshold of 100 births with a birth rate of 11 births per 1,000 population equates to a cutoff of 9,091 persons), our sample changes only slightly. One percent of counties (0.13 percent of all births) are included in our current analysis but would be excluded from the population-based selection rule. On the other hand, 2.7 percent of counties and 0.28 percent of births would be included under a population-based selection rule but are excluded under our current selection criteria.

Since our analyses use data over time, it is important to characterize the significant changes in practice recommendations regarding C-sections occurring during this period. One major shift occurred starting in 1998, when the American College of Obstetricians and Gynecologists' (ACOG) bulletin yielded caution to administering a vaginal birth after a previous C-section. While this change is not directly related to our analysis of singleton first births, a second change occurred in 2001, when ACOG's bulletin began to recommend C-sections for a breech birth (Oster, 2018), following on the results of the Term Breech Trial (Hannah et al., 2000).

3.2 Infant Mortality

We supplement the detailed natality files with the linked birth and infant death files from the National Vital Statistics System. These data allow us to observe infant deaths up to one year after birth for all births described in Section 3.1. We focus on neonatal mortality (death during days 0-27 inclusive), as these early deaths are more closely connected to hospital interventions (Cutler and Meara, 2000). Because our data on infant and maternal morbidity are derived from the birth certificate, the prevalence of these conditions would be undercounted to the extent that complications arise after the birth certificate is completed. However, the linked infant death files enable us to measure all mortality in the neonatal period, even beyond discharge from the hospital.

3.3 Other Data

We use U.S. county population estimates from the National Institutes of Health (NIH) Surveillance, Epidemiology, and End Results Program (SEER). In addition, we use data on county demographics as well as the supply of healthcare facilities and practitioners from the Area Health Resource File (Griffith et al., 2021). Finally, we use data on medical malpractice liability payments at the state level from the National Practitioner Data Bank.

3.4 Defining Low- vs. High-Risk for a Poor Infant Health Outcome & C-Section Appropriateness

Our analysis necessitates distinguishing births by their underlying risk. We characterize a risk type in two different but related ways. First, we categorize mothers as either high- or low-risk of a poor infant health outcome. A birth is high-risk if it exhibits any of the following characteristics: preterm (less than 37 weeks gestation), maternal age under 18 or over 35, 19 or more prenatal visits, pregnancy-associated hypertension, maternal diabetes, eclampsia, or breech. We define a birth as low-risk if it exhibits none of these high-risk characteristics, generally following Card, Fenizia, and Silver (2023).[13]

Second, in later analyses, we model the probability of a C-section birth. Our predicted probabilities, which we label "C-section appropriateness," come from a model with maternal age, gestational age, prenatal visits, growth restrictions, breech, eclampsia, pre-eclampsia, and diabetes, in addition to several demographic and birth characteristics which have strong predictive power for C-section rates among singleton first births. This set of characteristics (excluding the demographic and birth characteristics) combines wherever possible those used in Card, Fenizia, and Silver (2023) with those used in Currie and MacLeod (2017). However, unlike Currie and MacLeod (2017), we only study first-birth singleton births, so previous birth history variables (i.e., previous C-section and parity) are excluded.

Our departure from Card, Fenizia, and Silver (2023) and Currie and MacLeod (2017) in our inclusion of various birth and demographic characteristics is motivated by our goal of understanding the role of place in the determination of C-section. Our ideal thought experiment to identify these place-based effects would involve random assignment of otherwise identical women to different counties within the United States. However, there is selective migration of individuals across counties in the United States,

[13] Departing from Card, Fenizia, and Silver (2023), we also include diabetes as a risk factor and exclude body mass index from our main analysis because it is only available for more recent years.

making the ideal thought experiment hard to emulate. To remove the part of the variation in place effects due to this selection and more closely resemble the thought experiment involving contrasts of otherwise identical mothers, we control for several covariates highly correlated with C-section.

Table 1 further details these distinctions. The column labeled "Risk factors categorizing high-risk births" lists the characteristics determining the high-risk/low-risk designation. The remaining four columns list the factors entering four different models of C-section appropriateness we estimate. When we control for the set of covariates listed in the "All covariates" column, we call this an "adjustment for all covariates," whereas an adjustment for the medical risks listed in the "Medical only" column we call an "adjustment for medical covariates." We contrast our selected characteristics directly in Appendix Table A.1 with those used by Card, Fenizia, and Silver (2023) and Currie and MacLeod (2017). As our sample period covers an extended period, we are unable to include characteristics available only in older data (e.g., cord prolapse). The R^2's of the four models are quite similar, all between 12 and 15 percent —implying that the additional variables beyond the medical risk variables provide little extra explanatory power for C-sections. However, as we will see, these additional variables spread out our distribution of predicted C-section appropriateness.

To assess how these two classifications compare quantitatively, Table 2 displays the cross-tabulations for the 1989-1991 and the 2015-2017 periods separately. In light of the fact that C-section appropriateness is a continuous measure, we create a dichotomous measure above and below 0.6. Virtually all births with high predicted appropriateness for C-section fall into our high-risk birth classification. About two-thirds of births with predicted appropriateness below 0.6 are classified as low-risk. These patterns are similar if we use only medical covariates to predict appropriateness (see Appendix Table C.1).

3.5 Descriptive Statistics

Table 3 presents summary statistics for our sample of natality-data counties. For the natality data, we calculate these statistics separately for two periods of time (1989-1991 and 2015-2017) bookending the sample period. Much of our analysis will focus on contrasts between these two periods, with the earlier period characterized by relatively low C-section rates, and the latter period characterized by relatively high C-section rates, at the national level.

Our sample from the natality data covers a balanced panel of 2,346 counties once we drop counties experiencing few births. This sample covers 75 percent of all counties in the United States and 98.6 percent of births. Counties vary widely in their population size and number of births (counties in the 75$^{\text{th}}$ percentile deliver four times as many births as those in the 25$^{\text{th}}$), and thus in our analysis we weight counties by number of births as appropriate. The share of births that are singleton first births (the primary focus of our analysis sample) is 32.7 percent (1989-1991) and 29.4 percent (2015-2017) with roughly one-third of these births categorized as high-risk and nearly all of the rest as low-risk.[14] Overall, the difference in C-section rates between the 25$^{\text{th}}$ percentile and the 75$^{\text{th}}$ percentile is nearly 8 percentage points for both periods (between a 32 to 39 percent difference). In percentage-point terms, the variation is larger amongst high-risk singleton first births than amongst low-risk singleton first births.

Looking across the two time periods and all births (including higher order births), one of the most stark patterns is the change in the C-section rate over time —rising from 0.24 in the earlier period to 0.32 in the later period. Interestingly, however, the standard deviation of the C-section rate across counties has stayed stable. Contrary to what one would have predicted given the increase in C-sections, the fraction of births that are high-risk singleton first births has dropped slightly in the most recent years.[15]

[14] A small portion of singleton first births have no observed high-risk characteristics but are missing data, and thus are not classified as either high or low-risk.

[15] Another pattern worth mentioning is the fall in infant morbidity. However, as the underlying measures for this composite measure change over time, it is not possible to know whether infant morbidity is

4 Results

4.1 Time Trends in C-Section Usage and Geographic Variation

Figure 1 documents the trend in county-level C-section rates over time, including the mean, median, and the interquartile range. Overall C-section rates for all births in the late 1980s and early 1990s hovered at 22 percent, rose rapidly between 2000 and 2010, and then flattened to their current level of 32 percent. The variation in C-section rates across counties has not changed appreciably over time as seen earlier in Table 3.

A naive but incorrect interpretation of the patterns in Figure 1 is that the rise in C-sections is due to a decline in vaginal births after a previous C-section (the practice recommendation came out in 1998). However, the time trend for singleton first births, the focal sample of our later analysis, mirrors that of the overall trend. Contrasting high-risk and low-risk singleton first births in Figure 2, the time trends share the same common shape (i.e., falling slightly during the 1990s, a significant rise during the 2000s, and a leveling during the 2010s) with a few nuances. First, C-section rates among the low-risk group are roughly fifty percent smaller than that of the high-risk group. Second, C-section rates for low-risk mothers have been declining during the past decade, whereas for high-risk mothers they continue to climb albeit at a slower rate than that experienced during the 2000s. Lastly, as with all births combined, cross-county variation for both high- and low-risk singleton first births has remained relatively stable in magnitude.

4.2 Place Effects in C-Section Usage

Digging into the cross-sectional variation in these C-section rates, there are significant geographic differences in C-section for high-risk singleton first births as seen in Figure 3. The change in the shading of the map is quite dramatic moving from the 1989-1991

truly falling.

period to the 2015-2017 period, with rates of C-section among high-risk mothers rising significantly.

Looking at Figure 3, in the 1989-1991 period, the South exhibited the highest rates of C-section, while the lowest rates were in the Mountain Census region. At the end of the analysis period, the variation looks quite similar —indicative of persistence in these rates over time albeit with higher levels. Geographic patterns for high-risk singleton first births for 1989-1991 deviate from those for low-risk singleton first births (Figure 4) in that high rates of C-section are scattered throughout counties from North to Southeast of the Mississippi rather than almost exclusively in the South.

A question arising from Figures 3 and 4 is the degree to which the geographic patterns are stable over time. Our data span nearly 3 decades, facilitating a longer-term examination of the persistence of geographic practice variation than is feasible in most settings. We display the graphical evidence of persistence in Figure 5, a binned scatter plot of "all covariate-adjusted" C-section rates across time periods. Recall that we refer to "all covariates" model as including the controls listed in Table 1. We compare the county's C-section rate over a three-year period at the beginning of the sample period (1989-1991) to the rate at the end of the sample period (2015-2017). Averaging across three years and weighting by the number of births in each county mitigates some of the concern about small county-level samples introducing noise and thus, dampening the across-time correlation. Despite substantial changes in medical technologies and obstetrics practices over our study period, rates are persistent for both low-risk and high-risk singleton first births. The slope of the relationship is slightly stronger for low-risk singleton first births but qualitatively the slopes are similar across the risk types.[16] For high-risk singleton first births, counties with a 10 percentage point higher C-section rate in the 1989-1991 period experienced a 4.4 percentage point higher C-section rate in the 2015-2017 period.

[16]The figures look similar if we adjust for medical covariates only. The slopes of the relationships are slightly larger. See Appendix Figure C.1.

Another question is whether the place effects are correlated across the two risk types. Are places that perform more C-sections for high-risk mothers, for whom a C-section may be more appropriate, also conducting more C-sections for low-risk mothers? Using the same time periods as in Figure 5, Figure 6 displays a binned scatterplot of the relationship between C-section rates for high-risk first births and that for low-risk first births. The C-section rates are highly correlated across the two risk types; the slope is 0.85 for the 1989-1991 period and smaller at 0.68 for the 2015-2017 period.[17] High C-section counties for high-risk women are high C-section counties for low-risk women and vice versa, a pattern we delve into further in subsequent analyses.

If diagnostic skills were the prevailing determinant of treatment choices (Currie and MacLeod, 2017), then places with better diagnosticians would arguably have higher (lower) rates of intervention for the high-risk (low-risk) populations. While our patterns do not rule out diagnostic skill as an important factor, the findings from Figure 6 suggest a less dominant role for diagnostic skill. Rather, the data are more consistent with variation in C-section arising from a common force that leads to higher intervention rates for all patients, whether that be due to differential surgical skill or external factors. For example, providers in different areas could differ in their overall inclination for surgery, with providers in high C-section counties more inclined to use medical interventions, including C-section, regardless of risk type. The malpractice environment, either in terms of the cost of malpractice insurance or the likelihood of a successful malpractice lawsuit, could lead providers to do more C-section deliveries for both low- and high-risk mothers (Baicker, Buckles, and Chandra, 2006). In Currie and MacLeod (2017), by performing a high volume of C-section deliveries, providers may gain a comparative advantage in C-section deliveries relative to vaginal deliveries —resulting in higher rates of C-section for both high- and low-risk births.

[17]An adjustment for medical covariates only leads to slightly-altered slopes (0.88 and 0.72 respectively). See Appendix Figure C.2.

4.3 Are Higher Rates of C-Section Correlated With Infant and Maternal Morbidity and Infant Mortality?

Without data on outcomes, it is difficult to assess welfare. To this aim, we explore how county C-section rates vary with infant and maternal morbidity and neonatal mortality.[18]

4.3.1 Neonatal Mortality

In Figure 7, we present the extent to which the cross-county patterns in C-section rates correlate with neonatal mortality for both analysis periods and both risk types.[19] Across both time spans for high-risk singleton first births, the slope of the relationship is positive (i.e., places with higher C-section rates have higher rates of neonatal mortality). The slope is modest —for the 1989-1991 period, increasing the C-section rate among high-risk singleton first births by 1 standard deviation is associated with an additional 0.36 neonatal deaths per 1,000 births (a 3 percent increase). The relationship is flatter for the later period —falling by over 50 percent. In the bottom panel of Figure 7, we repeat this exercise except with low-risk singleton first births. Similar to the top panel figure for 1989-1991, there is an upward-sloping association but the slope is substantially smaller —6.5 percent the size of that for high-risk births for 1989-1991. The slope flips signs for 2015-2017 and is still considerably smaller than the analogous slope for high-risk mothers.[20] None of the estimated slopes are statistically different from 0 at

[18]We choose neonatal mortality, as opposed to infant mortality, as neonatal mortality is more closely connected to hospital interventions (Cutler and Meara, 2000).

[19]We have investigated this relationship, along with the relationship between C-section rates and infant and maternal morbidity, across three other subsamples, including births in counties offering continual hospital-based maternity ward services over the entirety of our study period, births to white mothers, and non-Medicaid births (only for the more recent period as Medicaid information was unavailable for the earlier time period). Qualitatively, the results look similar. In particular, higher rates of C-section are generally associated with the same or better health outcomes. However, for high-risk white mothers, the slope of the relationship between C-section and neonatal mortality is twice the size of that for the analysis sample. Specifically, the slopes for high-risk white mothers in the 1989-1991 and 2015-2017 time periods are 8.1 and 3.9, respectively.

[20]In a model adjusted for medical covariates only (see Appendix Figure C.3), one slope changes sign (i.e., the slope for high-risk mothers in the 1989-1991 period is negative) but the general conclusions hold

the 5 percent level. Collectively, these correlations imply a weak association between C-section rates and neonatal mortality, especially for low-risk births.

The correlations found in Baicker, Buckles, and Chandra (2006) differ somewhat from those of Figure 7. In their study of the 198 most populous counties for the years 1995-1998, they estimate a negative relationship between C-sections and neonatal mortality. A one standard deviation increase in the C-section rate is correlated with a drop in neonatal mortality of 0.02 per 1,000 births. However, none of these correlations are statistically different from zero, so the general conclusion from both our analysis and that of Baicker, Buckles, and Chandra (2006) is that there is a weak association between cross-sectional county variation in C-section use and county-level neonatal mortality rates.

4.3.2 Infant Morbidity

Figure 8 displays relationships between county C-section rates and our constructed measure of infant morbidity. Following Fischer, Royer, and White (Forthcoming), this morbidity measure is defined as the presence of any of the following conditions: moderate or heavy meconium staining, birth injury, seizure, or assisted ventilation. Note that due to changes in the birth certificate, the first two conditions (meconium and injury) are only available during the 1989-1991 period and thus infant morbidity indices are not directly comparable over time. The relationships in this figure are consistently downward-sloping with higher C-section rate counties having improved infant morbidity.[21] The associations are stronger than those for neonatal mortality. A 10 percentage point increase in the C-section rate is associated with changes in infant morbidity that are 12 to 18 percent of the mean. For neonatal mortality, the estimated effect of an equivalent rise in the C-section rate is 2.0 to 3.6 percent of the mean.

(i.e., weak associations between C-section rates and neonatal mortality).

[21] Looking at the components of infant morbidity separately, the county-level relationship between C-section rates and each infant morbidity component is downward-sloping.

For both low-risk and high-risk singleton births, the relationship is weaker for the later time period.[22] However, one should be cautious comparing these measures due to the mentioned changes over time in the index.

4.3.3 Maternal Morbidity

As a final set of health outcomes, we examine maternal morbidity in Figure 9. Following Fischer, Royer, and White (Forthcoming), maternal morbidity is defined by the presence of any of the following conditions: febrile, excessive bleeding, seizure, transfusion, third or fourth degree perineal laceration, ruptured uterus, unplanned hysterectomy, or admission to ICU.[23] Due to changes in the birth certificate, the first three conditions (febrile, excessive bleeding, and seizure) are only available during the 1989-1991 period, and the remaining five conditions are only available during the 2015-2017 period; as a result, maternal morbidity indices are not comparable over time. Here we do not consider maternal mortality due to issues of undercounting (MacDorman and Declercq, 2018) and its fortunate infrequency. However, if we expand our set of maternal health components to include maternal mortality, the shape of the relationship is consistent with those in Figure 9.[24]

The most apparent and consistent pattern across the four graphs in Figure 9 is the

[22] An adjustment for medical covariates only, instead of the full set of covariates, has minimal impact on our conclusions. See Appendix Figure C.4.

[23] Note that these measures exclude many complications occurring during the post-partum period, as these morbidity measures are collected when the birth certificate is completed (typically in a one or two days after the birth). Thus, complications such as ruptured sutures, postpartum hemorrhage, or infection are excluded from our maternal morbidity measure.

[24] However, this analysis treats each outcome equivalently, which may be inappropriate since a maternal death should arguably have more weight given its severity.

downward-sloping relationship.[25,26] This relationship is somewhat stronger than that for infant morbidity. A 10 percentage point increase in the C-section rate is associated with changes in maternal morbidity that are 14 to 32 percent of the mean. Taken together, the morbidity measures point to health-improving correlational returns to higher C-section rates, whereas the association of a county's C-section rate and neonatal mortality rates is weaker —exhibiting both positive and negative, albeit small, correlations.

4.4 Appropriateness of C-Section and Its Use Across Geographies

We next delve further into these county-level aggregate statistics by examining the relationship between C-sections and health outcomes through the lens of the appropriateness of care. Chandra and Staiger (2020) propose testing for allocative inefficiency by examining whether the probability of treatment varies across geographic space for patients of similar predicted probabilities of treatment, while the model in Currie and MacLeod (2017) suggests that procedural skill could increase C-section rates across the distribution of patient appropriateness. In essence, we want to know how a particular mother's probability of a C-section varies across counties when the mother moves from a high C-section county to a low C-section county or vice versa. Of course, it is impossi-

[25] The slopes are qualitatively similar if we control for the medical covariates rather than the full set of covariates. See Appendix Figure C.5. In an unreported analysis available upon request, we compare the magnitude of our results to those in Card, Fenizia, and Silver (2023) for the one overlapping measure of maternal morbidity between our studies: perineal laceration. Our results are directionally similar, though substantially smaller in magnitude. Card, Fenizia, and Silver (2023) find that delivery at a high C-section hospital reduces the incidence of laceration by 7 percentage points, while we find that delivering in a high C-section county is associated with a 0.5 percentage point lower rate of laceration. The sources of variation in each study are distinct though —the present study using cross-county comparisons and Card, Fenizia, and Silver (2023) utilizing a causal framework leveraging relative distance to a high C-section vs. a low C-section hospital.

[26] If we examine the maternal morbidity components separately, most of the estimated slopes are negative with a few exceptions. This includes maternal seizure (both high- and low-risk births) during the 1989-1991, transfusion for high-risk births during the 2015-2017 period, and ruptured uterus for high-risk births during the 2015-2017 period. For first births, a ruptured uterus would likely precipitate a C-section (i.e., not be caused by a C-section). None of the estimated slopes from these positively-sloped relationships, however, are statistically significant. Collectively we feel that these additional analyses, with a few exceptions, point to similar conclusions based on our maternal morbidity index measure.

ble to observe a mother delivering in two places at once. The feasible comparison is to contrast women with similar C-section propensities (based on their observable characteristics aside from their place of residence) and see how their probability of C-section varies across counties differing in their propensity for C-section deliveries.

To do this type of analysis, we create a model of C-section appropriateness. We first obtain covariate-adjusted C-section rates for year t by regressing whether a birth is delivered via C-section on county fixed effects and a comprehensive set of individual-level characteristics for year t. We provide details on these characteristics in Section 3.1. From that regression, the covariate-adjusted county C-section rate is the estimated county fixed effect. The individual predicted probability of C-section ("C-section appropriateness") is, for each individual birth, the prediction using all of the individual covariates but excluding the county fixed effects. The intent of this measure of appropriateness is to remove the contribution of the county and only consider individual-level circumstances. Thus, individuals with the same individual-level covariates will have the same appropriateness even if they live in counties that differ in their underlying rate of C-sections. However, because of possible selection on unobservables (e.g., physicians sorting across geographies based on surgical skill, heterogeneous patient preferences across counties), we are cautious about interpreting the correlations we examine. Additionally, the simultaneous partialling-out of county fixed effects means that the risk model is identified using only within-county variation in patient characteristics, addressing the potential bias that could arise from high-risk mothers sorting to high-C-section counties.

Appendix Figure B.1 displays the distribution of our predicted rates of C-section based on a prediction model using all covariates for 2015-2017. The distribution is bunched around 0.25. Nearly all of the distribution has a predicted probability of C-section below 0.4. There is little mass between 0.4 and 0.8. As such, our statistical inference will be much less precise for probabilities outside of the range of 0.1 to 0.4. One benefit of using the full set of covariates for adjustment is that the predicted probability of C-section is smoother (see Appendix Figure C.6 for the distribution of C-section appropriateness estimated by the model with only medical covariates). In the medical covariate model, due to the few number of discrete medical covariates, mothers are bunched into fewer values.

This model of appropriateness and its usefulness for categorizing mothers based on their C-section probability rests on several assumptions worth discussing. First, our prediction model assumes the individual-level covariates are additively-separable risk factors for a C-section apart from geography.[27] Second, while the natality data cover many of the inherent risks of delivering vaginally, the set may be incomplete, especially if physicians have access to additional patient information that influences their decision making. Third, this model assumes that appropriateness can be constructed from treatment choices made in the aggregate. To the extent that this model "learns from the crowd," if the crowd is wrong, the model will be as well. These concerns notwithstanding, our C-section appropriateness score is strongly related to the probability of C-section, as shown below, suggesting it provides a useful ranking of patients for analyzing the allocation of C-sections across and within counties.[28] Finally, our inclusion of preterm birth (defined as gestation $<$ 37 weeks) as a predictor in these models could be problematic if many C-sections are scheduled before the 37-week cutoff (a practice termed "early elective delivery"). To the extent that our indicator for preterm birth is actually an outcome of the decision to perform a C-section, we will mischaracterize those births as "appropriate" for C-section. While this is not common practice in today's environment, it was more prominent in the 1990s and early 2000s (Buckles and Guldi, 2017).

[27]A fully interacted model would relax this assumption, but the large number of covariates considered would raise concerns about the stability of estimates given small cell sizes.

[28]Related work by Currie, MacLeod, and Van Parys (2016) has addressed this issue by estimating patient appropriateness using only the decisions of better-trained providers. We do not have details on individual-level providers, but we have explored how our appropriateness measure varies with the sample used to estimate the appropriateness model (i.e., we estimate the appropriateness model using only the top 25 percent of counties in terms of the lowest infant mortality rates). The correlation of our preferred C-section appropriateness measure and this modified measure is 0.996 —indicating some insensitivity of the prediction model to the sample used to estimate appropriateness.

As a result, this issue could affect our findings from the 1989-1991 period but is of less concern for the 2015-2017 period (as the two periods use separate prediction models).

4.4.1 C-Section Appropriateness and the Use of C-Section

Figure 10 plots the measure of individual-level C-section appropriateness measured in percentage points against the probability of C-section for three geographic groupings (low-rate, medium-rate, and high-rate C-section counties).[29,30,31] Moreover, the consistency of the shape of the curve across the three groups indicates that these three types of counties respond similarly to the underlying patient-level characteristics in their determination of whether to perform a C-section.

Another notable feature of this figure is that the curves do not overlap or cross —meaning that higher C-section areas perform more C-sections across the entire support of predicted C-section risk probabilities. If worse diagnostic skills in high C-section areas drive the observed C-section patterns, the curve for high C-section counties should be flatter than that for lower C-section counties.

That is, under a model of diagnostic skill delivering geographic variation in C-section usage, the gap in usage across counties arranged by their C-section rates should be larger for those least appropriate for C-section than the gap across counties for those most appropriate for C-section. Instead, the geographic variation is more consistent with what Currie and MacLeod (2017) term "surgical skill." Physicians in areas with elevated C-section rates may have a comparative advantage in the surgical procedure

[29]The C-section rate is adjusted for all covariates and is measured for singleton first births.

[30]In honor of David Card, to preclude his frequent questioning of where the 45 degree line is, we include the 45 degree line. The raw C-section rate increases with the individual predicted C-section rate, indicating that rates of C-section are responsive to our underlying measure of appropriateness. This finding is true for all three groups across both time periods (1989-1991 in the top panel and 2015-2017 in the bottom panel).

[31]Considering a model adjusting for medical covariates only, the relationship between predicted probability of a C-section and the rate of C-section is similarly upward sloping and tracks the 45 degree line well. However, not surprisingly due to the lumpiness of the data, the patterns are less smooth. Also, for high levels of C-section appropriateness, C-section rates consistently fall below the 45 degree line —indicating that our prediction model overpredicts C-section rates. See Appendix Figure C.7.

(or at least believe they have such an advantage, as discussed in Chandra and Staiger (2020)), leading them to do C-sections more frequently across the C-section risk distribution. Alternatively, physicians in these places may have weaker risk tolerances or face stronger malpractice environments, possibly leading to higher rates of intervention without any skill differences, as suggested in Baicker, Buckles, and Chandra (2006). One other caveat to consider is that high C-section counties may be more sensitive to factors outside of our risk models that are present across the distribution of observed risk, such as fetal heart rate anomalies. In this case, our observed C-section usage patterns may be less sensitive to underlying risk than the true response function and thus, lead us to discount the diagnostic role in explaining geographic differences. However, it would be surprising to see such large and uniform differences across our appropriateness measure (a pattern not consistent with diagnostic-skill explanations) alongside strong patterns of diagnostic skill in terms of unobserved risks.

Also evident in Figure 10 is that the majority of the appropriateness distribution falls between 0.1 and 0.5, as each marker in the figure represents a percentile bin. This pattern has not changed much between the two time periods. Only 4 to 5 percent of the sample has very high rates of predicted C-section appropriateness (completely accounted for by the presence of a breech birth), a percent that has increased slightly over the two time periods. At these percentiles, the rate of C-section is very high —0.8 or above. As the rest of the paper relies on analysis by C-section appropriateness, it is important to keep in mind where the marginal C-sections lie in this distribution. Twenty percent of mothers have a predicted C-section appropriateness exceeding 0.33, whereas 30 percent of mothers have predicted C-section appropriateness above 0.29 (see Appendix Figure B.1). A useful benchmark is the *Healthy People 2030* target of a 23.6 percent C-section rate for low-risk first births. Thus, taking into consideration that any reasonable C-section target for higher-risk first births would be higher, the marginal births are likely to lie between C-section appropriateness levels of 0.29 and 0.33.

4.4.2 C-Section Appropriateness and Health Outcomes

To see how these practice patterns translate into outcomes, Figure 11 plots unadjusted rates of neonatal mortality as a function of C-section appropriateness (the same x-axis as in Figure 10).

With the exception of the highest and lowest percentiles of C-section appropriateness in the 1989-1991 period, rates of neonatal mortality are generally increasing with respect to appropriateness.[32] This finding means that patients for whom a C-section may be more appropriate generally experience higher risks of neonatal mortality. Across much of the support of appropriateness (i.e., between 0.1 and 0.4), rates of neonatal mortality are low and not visually distinct across low, medium, and high C-section counties. At higher levels of appropriateness, neonatal mortality rates move together across the three types of counties with none of the low, medium, or high C-section counties consistently experiencing lower rates of neonatal mortality. From a statistical perspective, distinctions between the three county types is difficult with neonatal mortality because of its low incidence. Only for levels of appropriateness exceeding 0.5 do we see neonatal mortality rise appreciably. This result serves as a good motivation for examining other outcomes such as infant morbidity and maternal morbidity which have higher incidence rates.

In Figure 12, we follow the same layout as Figure 11 but instead use maternal morbidity and infant morbidity as our outcomes. The shape of these relationships is quite distinct from Figure 11. In particular, for three of the four graphs (the exception is infant morbidity for 2015-2017), singleton first births highly appropriate for C-section (with an appropriateness index exceeding 0.8) exhibit nearly the lowest rates of maternal or infant morbidity across the appropriateness distribution. It is plausible, however not testable with our data, that there is a lot less ambiguity about the usefulness of C-section for these births, and because these mothers receive appropriate care, they experience

[32] We have investigated the odd spike in the 1989-1991 figure; it is due to the fraction of breech births discretely changing starting with this percentile.

better outcomes.

Despite having the highest rates of C-section, high C-section counties do not have the highest incidence of maternal morbidity and infant morbidity. For 1989-1991, rates of maternal morbidity and infant morbidity are the lowest amongst high C-section counties across all levels of appropriateness. In the later period, high C-section counties exhibit equal or lower rates of infant and maternal morbidity for most of the support of appropriateness.

One caveat to interpreting the results in Figure 12 is that some of the conditions underlying the morbidity measures may be mechanically related to the delivery type. The only underlying measure that stands out in this respect is perineal laceration, which is very unlikely for women who deliver by C-section. However, this condition is not available during the early period (1989-1991), where we also observe an improvement in maternal morbidity for higher C-section counties, indicating that our results are not dependent on this condition alone. Nevertheless, because these potentially more mechanical measures are related to welfare, we include them in our analysis.

Taking Figures 11 and 12 together, it appears that high C-section counties generate better health outcomes, or at the very least do not appear to produce *worse* outcomes, a pattern that holds throughout the appropriateness distribution. These conclusions are generally robust to only adjusting for medical covariates (see Appendix Figures C.8 and C.9). Of course, the outcomes we study are limited. Specifically, a C-section delivery may affect the probability of breastfeeding, future pregnancy issues such as those relating to the placenta, a mother's long-term physical health, the infant's ability to breathe, etc., which are not covered here.[33] Setting aside these important limitations, one simple view of our findings is that the additional C-sections performed in high C-section counties are actually health-improving: morbidity rates are significantly lower in counties with higher C-section rates, while infant mortality rates are similar.[34]

[33]Source: `https://www.nhs.uk/conditions/caesarean-section/risks/`.

[34]Other studies, including Currie and MacLeod (2017) and Johnson and Rehavi (2016), conclude that

4.4.3 Characterizing Low, Medium, and High C-Section Counties

We adjust for individual covariates to remove the part of the variation due to the observable selection of people across geographic space. However, it is possible that low, medium and high C-section counties differ along other dimensions not yet explored, explaining why high C-section counties do not generally experience worse outcomes. To investigate selection in more detail, we contrast these counties in Tables 4a and 4b. In Table 4a, we investigate disparities across these three groups of counties based on data available in the natality files. In Table 4b, we instead use data from the Area Health Resource File and the National Practitioner Data Bank to contrast counties.

In Table 4a, counties with the highest C-section rates have fewer births than medium and low C-section counties. However, with the exception of preterm birth (gestational age < 37 weeks) where high C-section counties have elevated rates, the underlying risk factors for a C-section are similar. High C-section counties are more economically disadvantaged (higher fraction of Medicaid births) and have a higher fraction of Black residents than do counties with lower C-section rates.

On economic dimensions, high C-section counties, relative to low and medium C-section counties, have depressed income per capita and higher rates of unemployment and poverty as shown in Table 4b. These counties also have a higher share of minority residents and fewer college graduates than low C-section counties. Collectively, area-level omitted variables bias arising from the economic and demographic factors of high C-section counties is correlated with worse outcomes for their mothers and their babies. Controlling for such selection would likely lead to even better comparative outcomes for high C-section counties.

Health care resources also differ by the three county types. However, it is less clear

there is overuse, though Currie and MacLeod (2017) find that this overuse is only for low-risk mothers, while there appears to be underuse for high-risk mothers, suggesting reallocation of C-sections towards high-risk mothers and away from low-risk mothers would improve outcomes. Of course, a full accounting of the total welfare analysis should take into account the health consequences plus the resource cost (e.g., higher health care costs – Podulka, Stranges, and Steiner (2011); Corry, Delbanco, and Miller (2013)), which are unaccounted for here.

the direction of selection as the effects of the resources on outcomes are more ambiguous. Hospital resources, measured in terms of hospitals per birth and newborn beds/bassinets, are more abundant in high C-section counties. In contrast, the per-capita number of physicians and obstetricians/gynecologists is lower in high C-section counties. Per-capita malpractice payments are also higher in the high C-section counties. On net, the answer to why and if the high C-section counties experience better outcomes is not simply explained away by selection, at least on the basis of the characteristics in Tables 4a & 4b.

4.5 Appropriateness of C-Section and Its Use Across Racial Groups

Racial disparities in C-section use are significant. In 2017, the C-section rate for Black mothers was 35 percent whereas for white mothers it was 31 percent (Yang and Mullen, 2022). This gap has grown over the last decade. A key question is the degree to which Black and white mothers with similarly-measured C-section appropriateness have different propensities of having a C-section. Akin to the earlier figures, we plot our appropriateness measure against the actual C-section rate in Figure 13 (a).[35] We exclude race from the prediction model and do not consider variation across counties by their underlying C-section rates because of concerns about sample size. We display the distribution of individual predicted probabilities in Appendix Figure B.2, divided by race. The distributions in panel (a) resemble the overall pattern. Looking across the two racial groups, non-Hispanic Black mothers' distribution has more density at smaller values of appropriateness. This pattern holds when body mass index (BMI) is added to the prediction model in panel (b), but for both groups, the inclusion of BMI spreads out the distribution —placing more density between 0.4 and 0.8.

Our first finding in Figure 13 is that, at the highest levels of appropriateness, non-Hispanic Black mothers have lower rates of C-section than non-Hispanic white moth-

[35]We display this distribution based on adjustment using only medical covariates in Appendix Figure C.6.

ers. The pattern flips for predicted C-section rates below 0.6. Using 2016 natality data along with the 10-point scale of the Robson Ten-Group Classification System, Valdes (2021) also documents higher rates of C-section delivery among Blacks at the high end of the "C-section risk" distribution and lower rates of C-section delivery at the lower end of the "C-section risk" distribution. More generally, our finding for Black mothers is consistent with the notion that minorities experience less tailored treatment choices, a finding that complements those in Johnson and Rehavi (2016) and Card, Fenizia, and Silver (2023), among others.

In the remaining panels of Figure 13, we examine how these C-section differences relate to outcomes. Across the spectrum of C-section appropriateness, the rates of neonatal mortality for non-Hispanic Black mothers exceed that of non-Hispanic white mothers (panel (b)).[36] The gap is largest for births we estimate to be most appropriate for C-section. The patterns for infant morbidity resemble that for neonatal mortality, but the gap emerges at a higher level of appropriateness (0.5 for infant morbidity and 0.3 for neonatal mortality). Like the infant health outcomes, maternal morbidity among Black mothers at high levels of C-section exceeds that for white mothers, but unlike the infant health outcomes, maternal morbidity for Black mothers is lower than that for white mothers at predicted probabilities of C-section below 0.4. The sizable elevated risk of morbidity and mortality for Black mothers at the high end of the appropriateness distribution holds when the C-section prediction model includes only medical risks (Appendix Figure C.10) or additionally includes body mass index (Appendix Figure D.1), a factor mentioned as a contributing factor in the rise in C-section use (Chu et al., 2007). Taken literally and ignoring the issues of patient and provider selection, these results are suggestive that the rate of C-section is too low for Black mothers most appropriate for C-section. High-risk Black mothers and their infants might benefit from higher C-section rates. This particular form of potential underuse among Black patients mirrors

[36]This racial gap in neonatal mortality also appears when contrasting mothers of similar incomes (Kennedy-Moulton et al., 2022).

findings from the literature (e.g., Abaluck et al., 2016).

5 Discussion and Conclusion

In this paper, we use data on births and infant deaths covering the entirety of the United States (1989-2017) to study county-level variation in the use of C-section, the most common surgical procedure. We document how county-level C-section rates have recently leveled off after a dramatic rise in the 2000s and have had strong persistence over the three decades. Looking through a lens of appropriateness of care, we detail how C-section rates in high C-section counties are higher than those in counties with lower C-section rates, both for high-risk and low-risk mothers. By many measures available in our data, health outcomes are *better* in high C-section counties, a fact that contrasts with much of literature on overuse and "flat of the curve medicine" (e.g., Fuchs, 2004).

This observation that high C-section counties exhibit higher rates of C-section along the continuum of C-section appropriateness is more consistent with surgical skills, rather than diagnostic skills, driving the variation in C-section rates (Currie and MacLeod, 2017). Under a goal of reducing C-sections for low-risk mothers, our results imply potentially adverse consequences for high-risk mothers. These mothers might fare worse if the target leads to spillover effects onto higher C-section risk groups. In particular, the result of such a policy could be a fall in C-sections for mothers more appropriate for C-section or a drop in physicians' surgical performance through a deterioration of surgical skills.

Together these facts provide insights into the key question in the geographic variation literature: are these geographic differences attributable to allocative inefficiencies (Chandra and Staiger, 2020) with some counties performing too many C-sections and others too few? Our answer to this question is not straightforward. As best we can measure, the outcomes of high C-section counties would worsen under a national standardization of care with a policy target of X percent of C-sections, where X is below the current rate in high C-section counties. On the other hand, setting a target above the current C-section rate of low C-section areas could improve outcomes. But these conclusions are subject to the caveat that high C-section counties are different. However, adjustment for demographic and economic characteristics only amplifies the health-outcome advantage of high C-section counties, as these counties tend to be relatively disadvantaged.

One policy effort to standardize care, and thus potentially remove geographic disparities, is the use of decision aids to guide treatment decisions (e.g., $CHADS_2$ for atrial fibrillation, Abaluck et al. (2020)). However, the complexity of the C-section decision process does not lend itself well to the easy adoption of such decision aids. For example, in 2015, the World Health Organization adopted the Robson classification, a 10-group classification based on 5 maternal characteristics —parity, the presence of multiples, previous C-section, onset of labour, gestational age, and fetal presentation (e.g., breech is one presentation).[37] The Robson classification has some embedded complexities —the interaction of these 5 characteristics determine the grouping. Foreseeably, this classification is only used to standardize reporting of C-sections across risk groups rather than inform C-section decisions. The ambiguity and complexity of the environment instead leads to very simple decision aids allowing for little or no provider discretion, typically involving only one present underlying condition. For instance, two examples are "automatic" C-sections following a previous C-section birth or a breech presentation. With the development of more complicated algorithms for the allocation of care, our work suggests the environment of C-section is ripe for innovation.

While our paper focuses on county-level variation, it misses important within-area variation (Epstein and Nicholson, 2009). We abstract from the rich variation across individual hospitals and doctors highlighted in other studies (Epstein and Nicholson, 2009; Currie and MacLeod, 2017; Card, Fenizia, and Silver, 2023). A natural next step is to understand whether the documented observations of (1) persistence in C-section usage

[37] `https://www.who.int/publications/i/item/9789241513197`

over time and (2) the elevation of C-section rates across all risk propensities among higher C-section areas are replicated when the unit of observation is a hospital or a physician.

References

Abaluck, Jason, Leila Agha, David C Chan Jr, Daniel Singer, and Diana Zhu. 2020. "Fixing misallocation with guidelines: Awareness vs. adherence." NBER Working Paper No. 27467.

Abaluck, Jason, Leila Agha, Chris Kabrhel, Ali Raja, Arjun Venkatesh et al. 2016. "The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care." *American Economic Review* 106 (12):3730–3764.

Almond, Douglas and Janet Currie. 2011. "Killing me softly: The fetal origins hypothesis." *Journal of Economic Perspectives* 25 (3):153–72.

Baicker, Katherine, Kasey S Buckles, and Amitabh Chandra. 2006. "Geographic variation in the appropriate use of cesarean delivery: Do higher usage rates reflect medically inappropriate use of this procedure?" *Health Affairs* 25 (Suppl1):W355–W367.

Battaglia, Emily. Forthcoming. "The effect of hospital closures on maternal and infant health." *American Journal of Health Economics* .

Buckles, Kasey and Melanie Guldi. 2017. "Worth the wait? The effect of early term birth on maternal and infant health." *Journal of Policy Analysis and Management* 36 (4):748–772.

Card, David, Alessandra Fenizia, and David Silver. 2023. "The Health Impacts of Hospital Delivery Practices." *American Economic Journal: Economic Policy* 15 (2):42–81.

Chan, David C, Matthew Gentzkow, and Chuan Yu. 2022. "Selection with variation in diagnostic skill: Evidence from radiologists." *The Quarterly Journal of Economics* 137 (2):729–783.

Chandra, Amitabh and Douglas O Staiger. 2007. "Productivity spillovers in health care: Evidence from the treatment of heart attacks." *Journal of Political Economy* 115 (1):103–140.

———. 2020. "Identifying Sources of Inefficiency in Healthcare." *The Quarterly Journal of Economics* 135 (2):785–843.

Chu, Susan Y, SY Kim, Christopher H Schmid, Patricia M Dietz, WM Callaghan, J Lau, and KM Curtis. 2007. "Maternal obesity and risk of cesarean delivery: a meta-analysis." *Obesity Reviews* 8 (5):385–394.

Chyn, Eric and Na'ama Shenhav. 2022. "Place Effects and Geographic Inequality in Health at Birth." NBER Working Paper No. 30424.

Coleman, James, Elihu Katz, and Herbert Menzel. 1957. "The diffusion of an innovation among physicians." *Sociometry* 20 (4):253–270.

Cooper, Zack, Stuart Craig, Charles Gray, Martin Gaynor, and John Van Reenen. 2019. "Variation in health spending growth for the privately insured from 2007 to 2014." *Health Affairs* 38 (2):230–236.

Corry, Maureen P, Suzanne F Delbanco, and Harold D Miller. 2013. "The cost of having a baby in the United States." *Truven Health Analytics, Greenwood Village, CO, USA* .

Currie, Janet and W Bentley MacLeod. 2017. "Diagnosing expertise: Human capital, decision making, and performance among physicians." *Journal of Labor Economics* 35 (1):1–43.

Currie, Janet, W Bentley MacLeod, and Jessica Van Parys. 2016. "Provider practice style and patient health outcomes: The case of heart attacks." *Journal of Health Economics* 47:64–80.

Cutler, David, Angus Deaton, and Adriana Lleras-Muney. 2006. "The determinants of mortality." *Journal of Economic Perspectives* 20 (3):97–120.

Cutler, David M and Ellen Meara. 2000. "The technology of birth: Is it worth it?" In *Forum for Health Economics & Policy*, vol. 3. De Gruyter.

Deryugina, Tatyana and David Molitor. 2020. "Does when you die depend on where you live? Evidence from Hurricane Katrina." *American Economic Review* 110 (11):3602–3633.

Epstein, Andrew J and Sean Nicholson. 2009. "The formation and evolution of physician treatment styles: an application to cesarean sections." *Journal of Health Economics* 28 (6):1126–1140.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of geographic variation in health care: Evidence from patient migration." *The Quarterly Journal of Economics* 131 (4):1681–1726.

———. 2021. "Place-based drivers of mortality: Evidence from migration." *American Economic Review* 111 (8):2697–2735.

Fischer, Stefanie J, Heather Royer, and Corey D White. Forthcoming. "Health Care Centralization: The Health Impacts of Obstetric Unit Closures in the US." *American Economic Journal: Applied Economics* .

Fuchs, Victor R. 2004. "More variation in use of care, more flat-of-the-curve medicine: Why does it occur? What should be done about it?" *Health Affairs* 23 (Suppl2):VAR–104.

Glover, J Alison. 1938. "The incidence of tonsillectomy in school children."

Griffith, Kevin N, Yevgeniy Feyman, Samantha G Auty, Erika L Crable, and Timothy W Levengood. 2021. "County-level data on US opioid distributions, demographics, healthcare supply, and healthcare access." *Data in Brief* 35:106779.

Hannah, Mary E, Walter J Hannah, Sheila A Hewson, Ellen D Hodnett, Saroj Saigal, Andrew R Willan, and Term Breech Trial Collaborative. 2000. “Planned caesarean section versus planned vaginal birth for breech presentation at term: A randomised multicentre trial.” *The Lancet* 356 (9239):1375–1383.

Hyde, Matthew J, Alison Mostyn, Neena Modi, and Paul R Kemp. 2012. “The health implications of birth by Caesarean section.” *Biological Reviews* 87 (1):229–243.

Johnson, Erin M and M Marit Rehavi. 2016. “Physicians treating physicians: Information and incentives in childbirth.” *American Economic Journal: Economic Policy* 8 (1):115–41.

Keag, Oonagh E, Jane E Norman, and Sarah J Stock. 2018. “Long-term risks and benefits associated with cesarean delivery for mother, baby, and subsequent pregnancies: Systematic review and meta-analysis.” *PLoS medicine* 15 (1):e1002494.

Kennedy-Moulton, Kate, Sarah Miller, Petra Persson, Maya Rossin-Slater, Laura Wherry, and Gloria Aldana. 2022. “Maternal and Infant Health Inequality: New Evidence from Linked Administrative Data.” NBER Working Paper No. 30693.

Kozhimannil, Katy Backes, Michael R Law, and Beth A Virnig. 2013. “Cesarean delivery rates vary tenfold among US hospitals; reducing variation may address quality and cost issues.” *Health Affairs* 32 (3):527–535.

MacDorman, Marian F and Eugene Declercq. 2018. “The failure of United States maternal mortality reporting and its impact on women’s lives.” *Birth (Berkeley, Calif.)* 45 (2):105.

Molitor, David. 2018. “The evolution of physician practice styles: Evidence from cardiologist migration.” *American Economic Journal: Economic Policy* 10 (1):326–56.

Mullainathan, Sendhil and Ziad Obermeyer. 2022. “Diagnosing physician error: A machine learning approach to low-value health care.” *The Quarterly Journal of Economics* 137 (2):679–727.

Oster, Emily. 2018. “Expert behavior change in response to best practice changes: Evidence from obstetrics.” Working Paper.

Podulka, Jennifer, Elizabeth Stranges, and Claudia Steiner. 2011. “Hospitalizations related to childbirth, 2008.” *Heathcare Cost and Utilization Project Statistical Brief 110* .

Shah, Neel. 2017. “Keynote to American Ob-Gyns: C-section rates linked to hospital complexities.” `https://www.ariadnelabs.org/resources/articles/keynote-to-american-ob-gyns-c-section-rates-linked-to-hospital-complexities/`. [Accessed 30-May-2023].

Silver, David. 2021. “Haste or Waste? Peer pressure and productivity in the emergency department.” *The Review of Economic Studies* 88 (3):1385–1417.

Singh, Manasvini. 2021. "Heuristics in the delivery room." *Science* 374 (6565):324–329.

Trish, Erin, Samuel Valdez, Paul B Ginsburg, Samantha Randall, and Steven M Lieberman. 2023. "Substantial Growth In Medicare Advantage And Implications For Reform: Study examines Medicare Advantage growth." *Health Affairs* 42 (2):246–251.

Valdes, Elise G. 2021. "Examining cesarean delivery rates by race: a population-based analysis using the Robson Ten-Group Classification System." *Journal of Racial and Ethnic Health Disparities* 8 (4):844–851.

Wennberg, John and Alan Gittelsohn. 1973. "Small area variations in health care delivery: A population-based health information system can guide planning and regulatory decision-making." *Science* 182 (4117):1102–1108.

Yang, Yanchao and Maggie Mullen. 2022. "Black-White Disparities of Cesarean Section Rate in the US: Trends and Determinants." SSRN Working Paper 4204876.

Table 1: Defining Risk and Appropriateness

	Risk factors categorizing high-risk births	*Risk factors predicting C-section appropriateness* All covariates	Excluding race	Excl. race + BMI	Medical only
			Racial disparities	*Appendix D*	*Appendix C*
Maternal age	<18 or >35	5-year bins	5-year bins	5-year bins	5-year bins
Gestational age	$<$ 37 weeks	$<$ 37 weeks	$<$ 37 weeks	$<$ 37 weeks	$<$ 37 weeks
Prenatal visits	$\geq$ 19	$\geq$ 19	$\geq$ 19	$\geq$ 19	$\geq$ 19
Growth restrictions*	✓	✓	✓	✓	✓
Breech	✓	✓	✓	✓	✓
Eclampsia	✓	✓	✓	✓	✓
Pre-eclampsia	✓	✓	✓	✓	✓
Diabetes	✓	✓	✓	✓	✓
Maternal education	–	✓	✓	✓	–
Maternal birth country	–	✓	✓	✓	–
Birth month	–	✓	✓	✓	–
Birth day of week	–	✓	✓	✓	–
Sex of child	–	✓	✓	✓	–
Father's info. present	–	✓	✓	✓	–
Race-by-Hispanic origin	–	✓	–	–	–
Body mass index	–	–	–	cubic polynomial	–
Adj. R^2 1989-1991	–	0.1383	0.1375	–	0.1342
Adj. R^2 2015-2017	–	0.1286	0.1272	0.1483	0.1221

Notes: This table describes the factors used to define risk and appropriateness. The first column describes the factors used to categorize high-risk vs. low-risk births. The second through fifth columns describe the factors that enter four different prediction models of C-section appropriateness that we estimate. See Table A.1 for comparison with select literature. For each prediction model, the adjusted R^2 is from a regression of C-section on the listed covariates and county fixed effects among singleton first births. *Growth restrictions are defined as below the 5th percentile of birthweight for gestational age. These cutoffs come from `https://srhr.org/fetalgrowthcalculator`.

Table 2: Distribution of Births Across Low- and High-Risk by Predicted C-Section Appropriateness

(a) 1989-1991

	Predicted C-section appropriateness (using all covariates)		
	≤ 0.6	*> 0.6*	*Overall*
Low-risk births	63.33 %	0.00 %	60.74 %
High-risk births	33.27 %	100.00 %	35.99 %
Unknown risk births	3.40 %	0.00 %	3.27 %
	100%	100%	100%

(b) 2015-2017

	Predicted C-section appropriateness (using all covariates)		
	≤ 0.6	*> 0.6*	*Overall*
Low-risk births	63.68 %	0.00 %	60.41 %
High-risk births	34.01 %	99.99 %	37.40 %
Unknown risk births	2.31 %	0.01 %	2.19 %
	100%	100%	100%

Notes: This table shows the relationship between two different approaches for assessing risk of C-section: (1) the categorization of births as low- or high-risk, a binary assignment based on medical factors only, and (2) the predicted C-section appropriateness, a continuum of risk estimated using all covariates. See Table 1 for the specific covariates used in each model. For this table, the cutoff of ± 0.6 was chosen due to the bimodal distribution exhibited by the predicted C-section appropriateness (see Figure B.1). Only singleton first births are represented. A small portion of singleton first births have no observed high-risk characteristics but are missing data, and thus are not classified as either low- or high-risk.

Table 3: Descriptive Statistics

	1989-1991				2015-2017			
	mean	sd	p25	p75	mean	sd	p25	p75
# Counties	2,346				2,346			
Population	105,260	307,227	18,549	77,921	135,564	377,944	21,115	102,119
Births	1,739	6,078	265	1,169	1,651	4,823	237	1,177
C-section rate	0.235	0.052	0.201	0.266	0.319	0.057	0.281	0.354
Fraction of births that are singleton 1st	0.327	0.046	0.298	0.355	0.294	0.038	0.272	0.316
C-section rate for singleton 1st births	0.243	0.063	0.202	0.280	0.283	0.065	0.242	0.319
Fraction of births that are low-risk singleton 1st	0.195	0.037	0.172	0.219	0.177	0.031	0.159	0.196
C-section rate for low-risk singleton 1st births	0.208	0.071	0.162	0.247	0.226	0.074	0.179	0.266
Fraction of births that are high-risk singleton 1st	0.125	0.032	0.102	0.144	0.111	0.025	0.096	0.125
C-section rate for high-risk singleton 1st births	0.303	0.087	0.247	0.353	0.376	0.093	0.324	0.429
Neonatal mortality rate per 1,000 births	5.488	4.269	2.865	7.375	4.097	3.763	1.406	5.780
Infant morbidity rate	0.079	0.045	0.052	0.097	0.048	0.031	0.027	0.062
Maternal morbidity rate	0.016	0.013	0.008	0.021	0.014	0.011	0.007	0.018

Notes: This table includes summary statistics for our sample of counties. We use a balanced panel and exclude counties that ever have fewer than 100 births. A small portion of singleton first births have no observed high-risk characteristics but are missing data, and thus are not classified as either high- or low- risk.

Table 4: Characteristics of Low, Medium, and High C-Section Areas

(a) Statistics Compiled from Natality Data

	1989-1991			2015-2017		
	Low CS	Medium	High CS	Low CS	Medium	High CS
# Counties	666	654	1,026	668	537	1,141
# Singleton 1st births per year	625	682	421	570	753	345
C-section rate (adjusted for all covariates)	0.192	0.232	0.284	0.238	0.283	0.336
Predicted C-section appropriateness	0.240	0.237	0.231	0.291	0.287	0.279
Maternal age	24.3	23.9	23.3	26.7	26.4	25.5
Share maternal age $<$18 or > 35	0.134	0.139	0.150	0.103	0.103	0.101
Share gestational age $<$37 weeks	0.092	0.097	0.100	0.087	0.089	0.102
Share w/ any prenatal visits	0.985	0.984	0.984	0.987	0.987	0.982
Share prenatal visits $\geq$19	0.032	0.032	0.034	0.033	0.033	0.030
Share w/ growth restrictions	0.080	0.080	0.080	0.084	0.083	0.087
Share breech	0.041	0.040	0.039	0.053	0.048	0.044
Share w/ eclampsia	0.006	0.006	0.007	0.004	0.003	0.003
Share w/ pre-eclampsia	0.042	0.040	0.041	0.076	0.072	0.076
Share w/ diabetes	0.019	0.018	0.017	0.056	0.054	0.054
Share non-Hispanic white	0.681	0.668	0.611	0.573	0.517	0.555
Share non-Hispanic Black	0.133	0.154	0.141	0.119	0.137	0.150
Share Hispanic	0.128	0.140	0.210	0.197	0.247	0.231
Share occurring in the county of residence	0.781	0.768	0.761	0.760	0.742	0.660
Share Medicaid	.	.	.	0.319	0.382	0.425

Notes: This table shows the average county characteristics for counties with low, medium, and high C-section rates. Counties are separated using the C-section rate for singleton first births (adjusted for all covariates), into terciles weighted by the number of singleton first births.

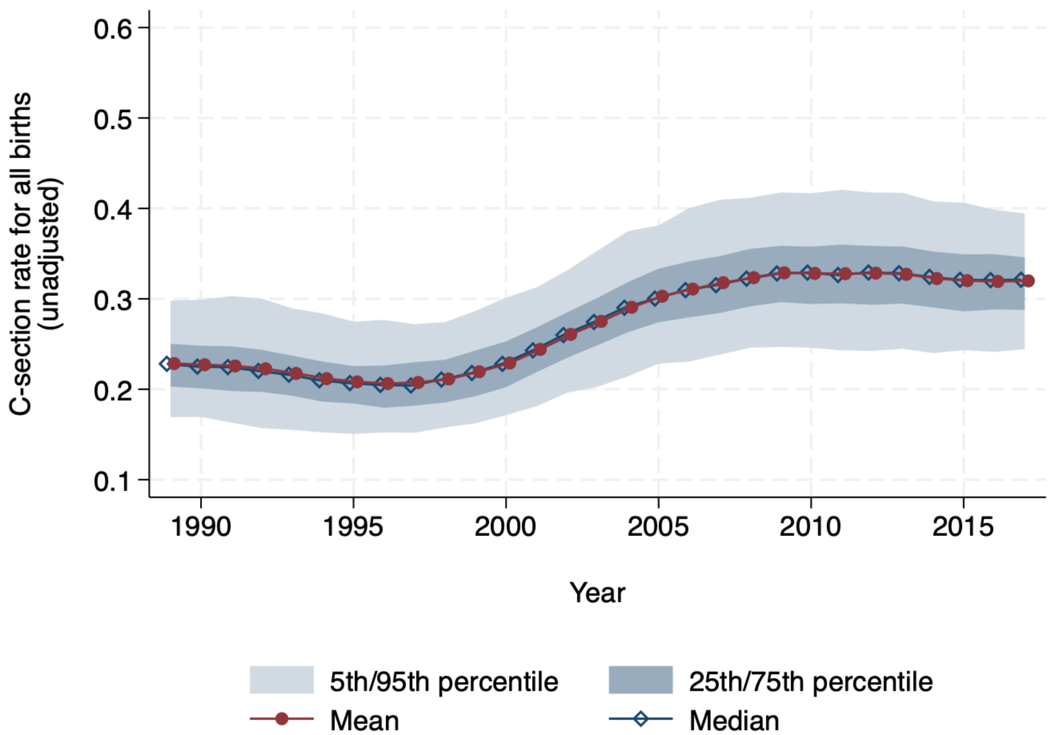
Table 4: Characteristics of Low, Medium, and High C-Section Areas (continued)

(b) Statistics Compiled from the Area Health Resource File and National Practitioner Data Bank

	1989-1991			2015-2017		
	Low CS	Medium	High CS	Low CS	Medium	High CS
Income per capita (2020 $)	38,709	37,978	34,415	59,285	54,206	49,532
Unemployment rate	0.055	0.058	0.067	0.046	0.049	0.052
Poverty rate	0.122	0.122	0.152	0.132	0.144	0.153
Share urban	0.818	0.797	0.734	0.861	0.879	0.773
Share w/ at least HS diploma	0.784	0.756	0.712	0.888	0.861	0.858
Share w/ 4+ years college	0.231	0.206	0.181	0.351	0.313	0.276
Hospitals per 1,000 births	1.489	1.447	1.699	1.395	1.281	1.705
Hospitals w/ med. school affiliation per 1,000 births	0.365	0.288	0.215	0.486	0.397	0.361
Hospital beds per 1,000 births	303	297	289	238	231	241
Newborn bassinets per 1,000 births	16.99	16.50	17.15	14.55	13.86	14.60
MDs per 1,000 births	153	130	106	292	231	192
OB/GYNs per 1,000 births	7.989	7.356	6.055	12.034	9.927	8.861
Medical residents per 1,000 births	0.547	0.444	0.467	0.013	0.020	0.015
Surgeons per 1,000 births	36.89	32.72	27.18	53.17	43.51	37.44
Operating rooms per 1,000 births	.	.	.	9.91	8.63	8.59
Malpractice # payments per 1,000 MDs	25.77	25.69	25.30	8.02	9.55	10.51
Obstetrics malpractice # payments per 1,000 OB/GYNs	44.63	43.15	43.54	11.43	14.04	14.11
Malpractice liability per MD (2020 $)	7,436	7,816	7,109	3,527	3,843	3,975
Obstetrics malpractice liability per OB/GYN (2020 $)	21,085	20,625	16,870	9,046	9,423	9,147

Notes: This table shows the average county characteristics for counties with low, medium, and high C-section rates. Counties are separated using the C-section rate for singleton first births (adjusted for all covariates), into terciles weighted by the number of singleton first births.

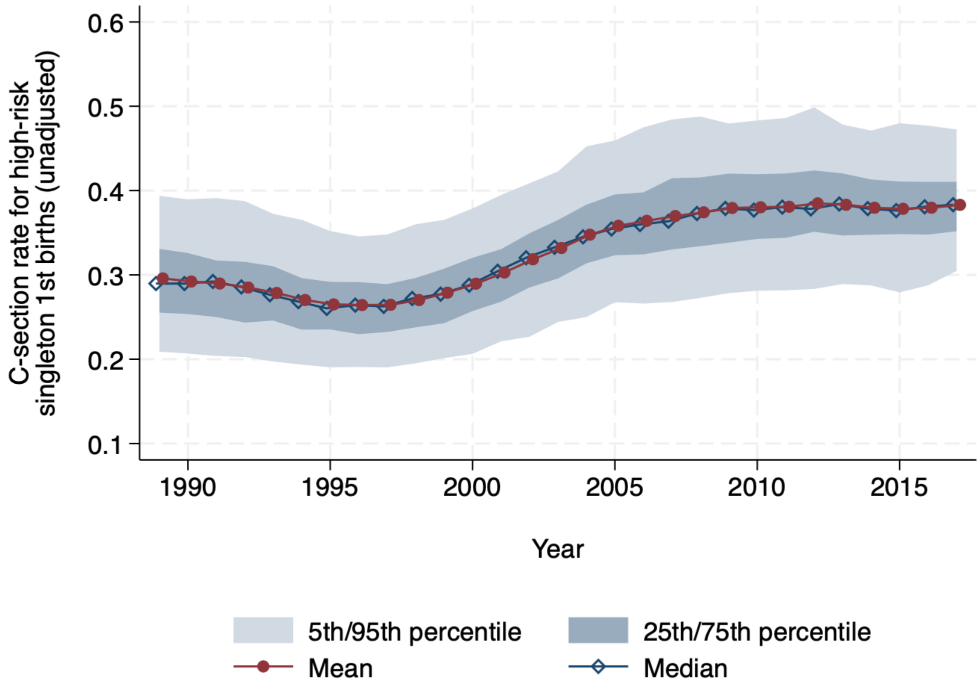
Figure 1: Time Trends in County C-Section Rates



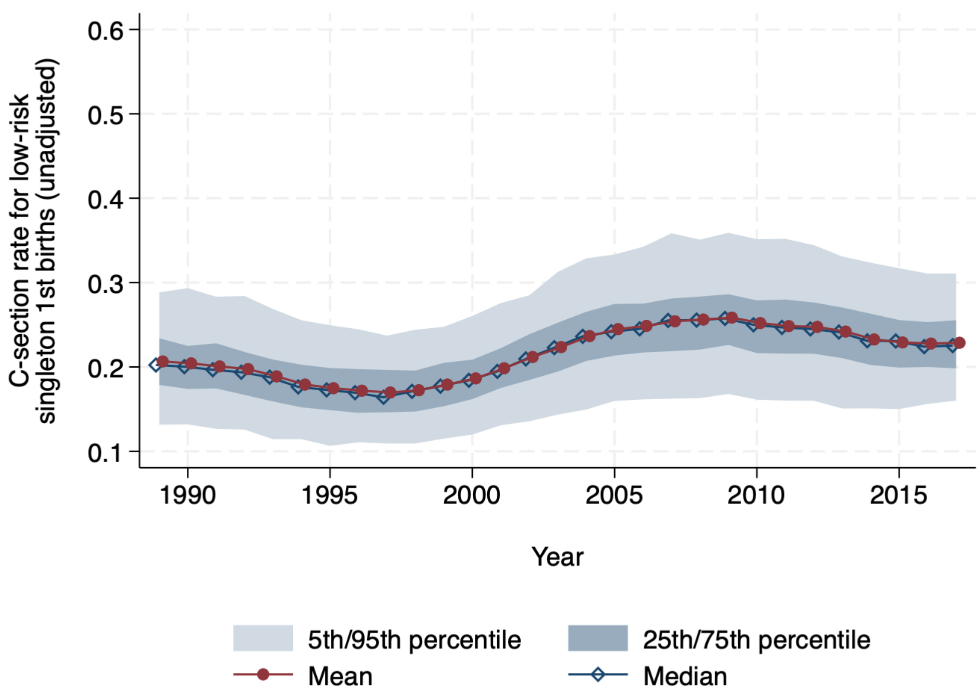
Notes: This figure shows the distribution of (raw) county C-section rates over time for all births (including higher order births). All statistics are weighted across counties by the number of births. The mean is the overall C-section rate, and 50 percent of births occur in counties with C-section rates in the interquartile range.

Figure 2: Time Trends in County C-Section Rates by Risk Group

(a) High-Risk Singleton 1st Births



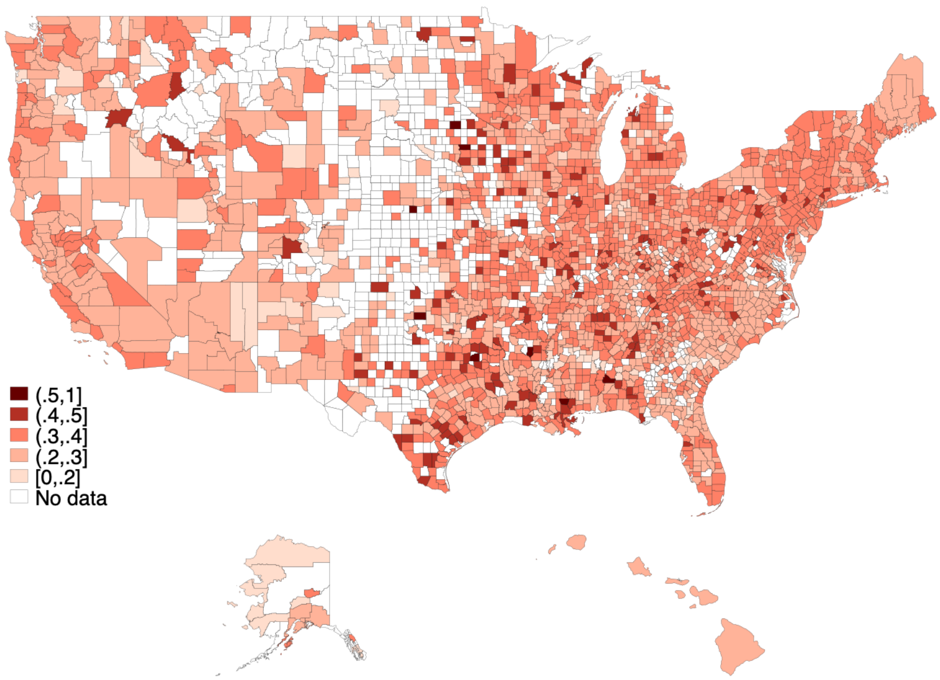
(b) Low-Risk Singleton 1st Births



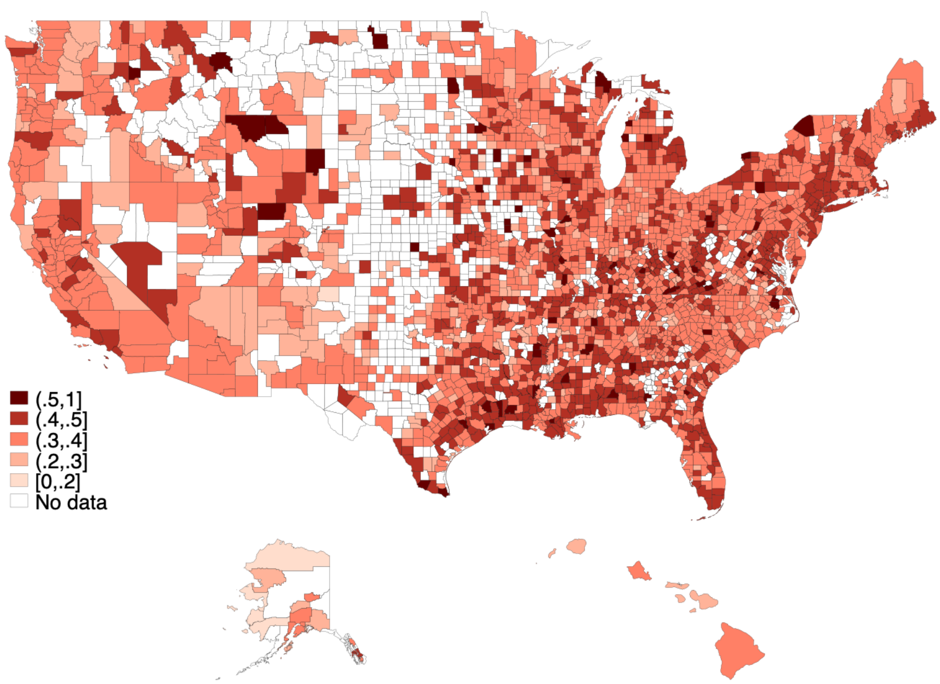
Notes: This figure shows the distribution of (raw) county C-section rates over time for low- and high-risk singleton first births. All statistics are weighted across counties by the number of relevant births.

Figure 3: Time Trends in County C-Section Rates Across U.S. for High-Risk Singleton 1st Births

(a) 1989-1991



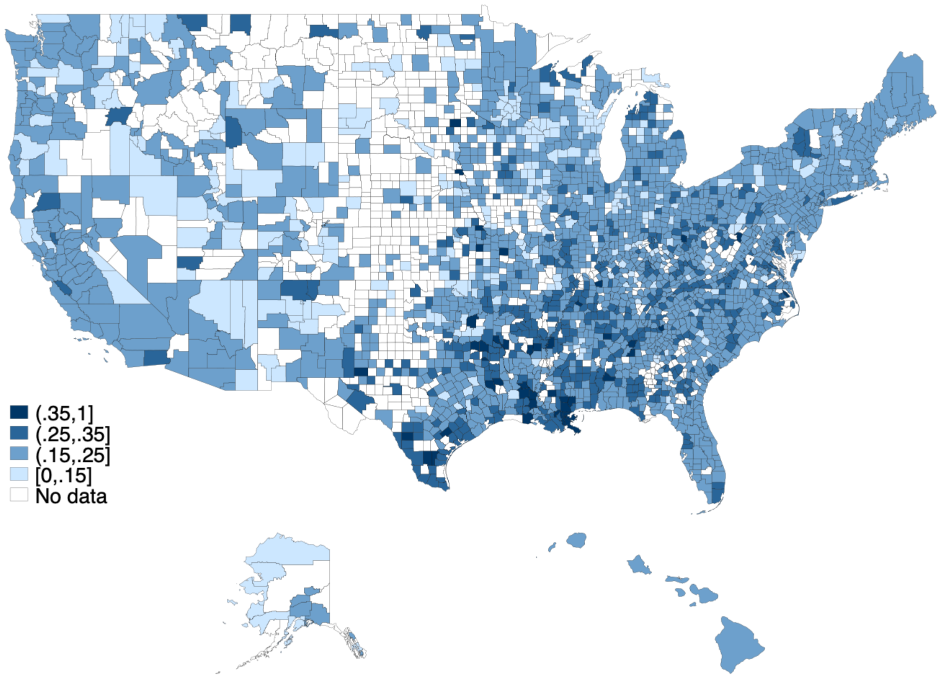
(b) 2015-2017



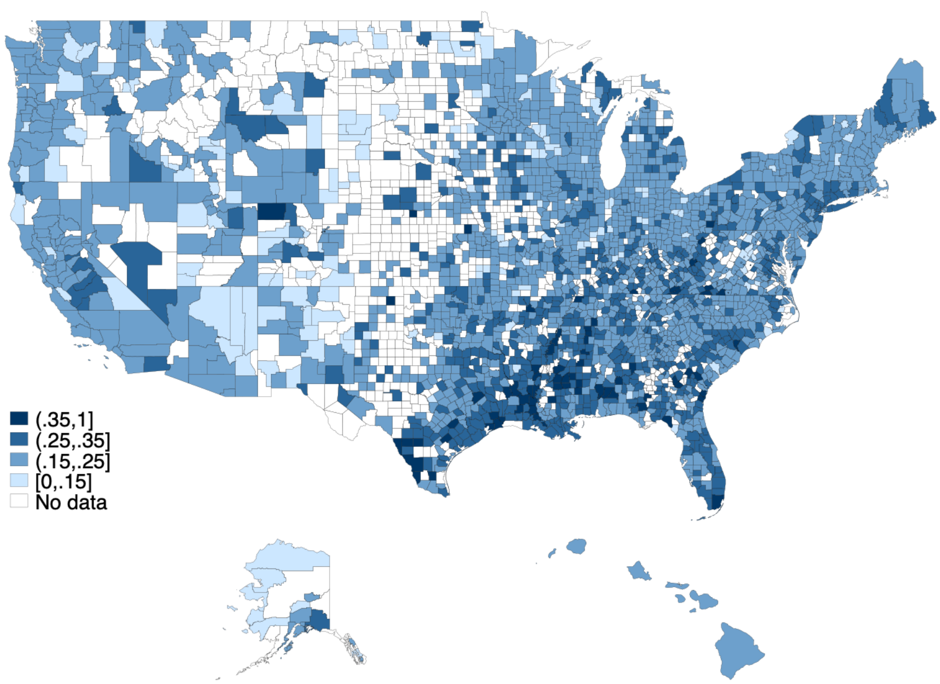
Notes: This figure shows the geographic distribution of (raw) C-section rates for high-risk singleton first births. Rates are the average within a county over the three-year period, weighted by the number of high-risk singleton first births in each year.

Figure 4: Time Trends in County C-Section Rates Across U.S. for Low-Risk Singleton 1st Births

(a) 1989-1991



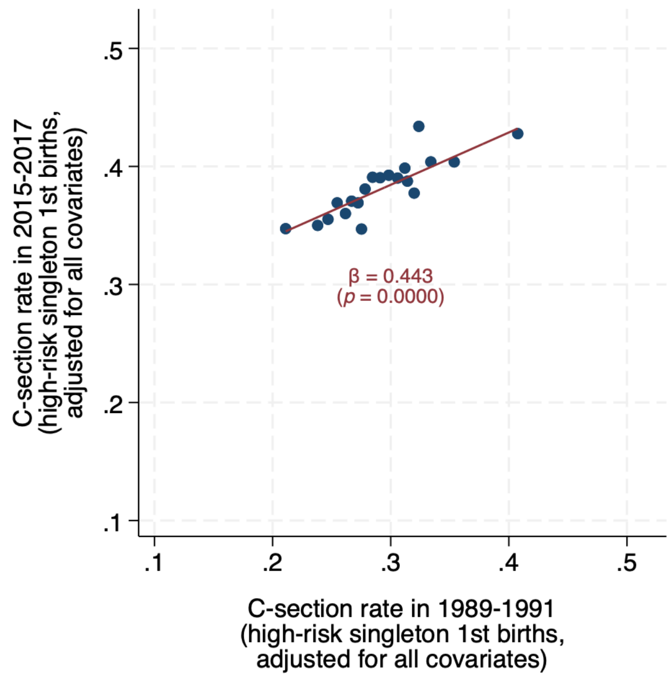
(b) 2015-2017



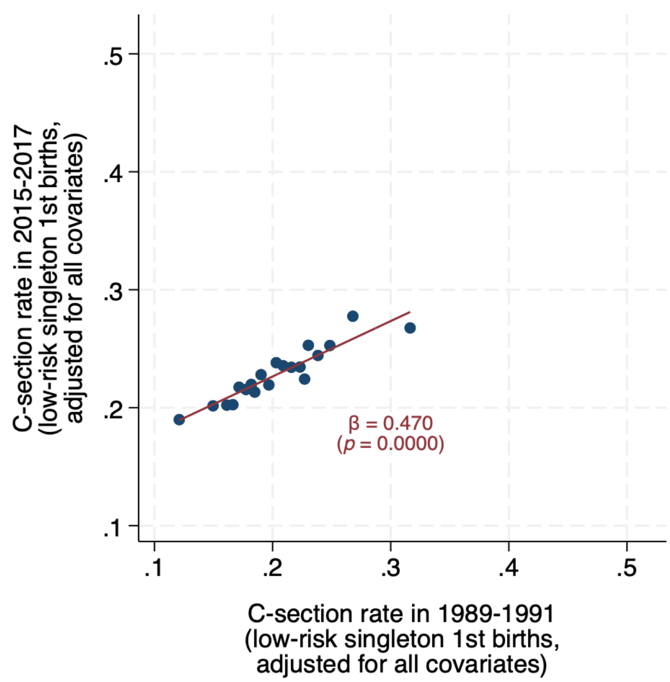
Notes: This figure shows the geographic distribution of (raw) C-section rates for low-risk singleton first births. Rates are the average within a county over the three-year period, weighted by the number of low-risk singleton first births in each year.

Figure 5: Persistence in Adjusted County C-Section Rates Over Time

(a) High-Risk Singleton 1st Births



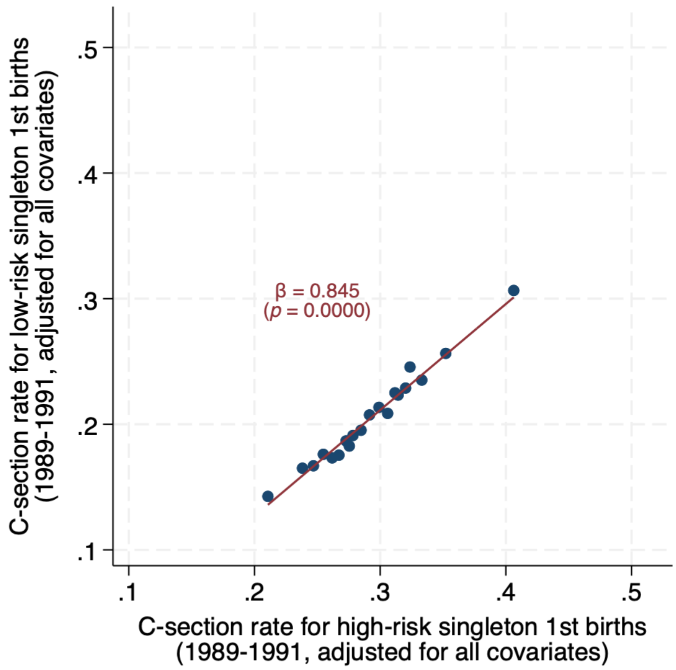
(b) Low-Risk Singleton 1st Births



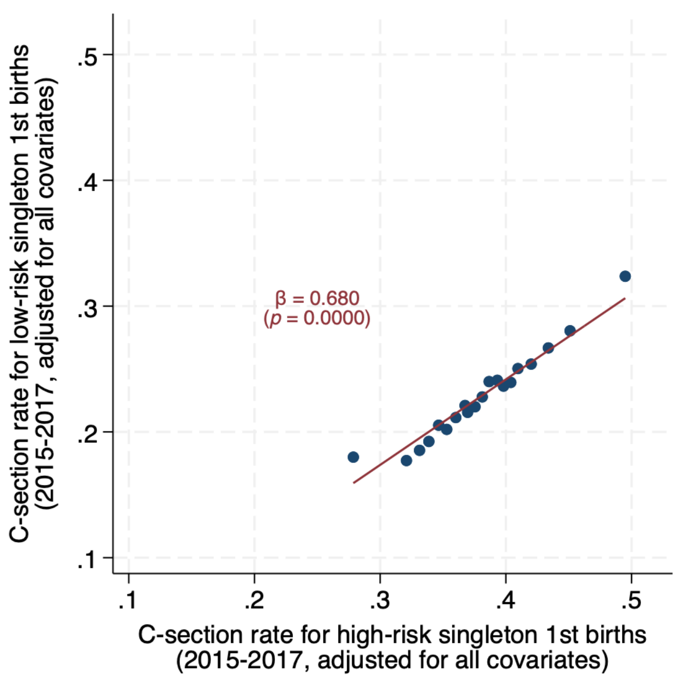
Notes: This figure shows binscatter plots of C-section rates across time periods. Linear fit and p-value are based on the underlying counties (prior to binning). C-section rates (adjusted for all covariates) are averaged over the three-year period, weighted by the number of relevant births in each year. Binscatter and linear fit are weighted by the number of relevant births in the county over all six years.

Figure 6: Correlation in Adjusted County C-Section Rates Across Risk Type

(a) 1989-1991



(b) 2015-2017

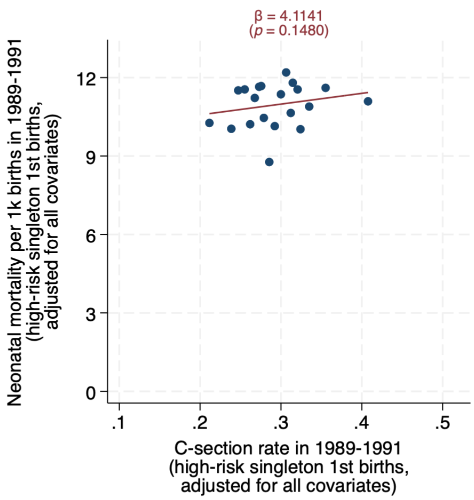


Notes: This figure shows binscatter plots of C-section rates across risk types. Linear fit and p-value are based on the underlying counties (prior to binning). C-section rates (adjusted for all covariates) are averaged over the three-year period, weighted by the number of relevant births in each year. Binscatter and linear fit are weighted by the number of singleton first births in the county over the three years.

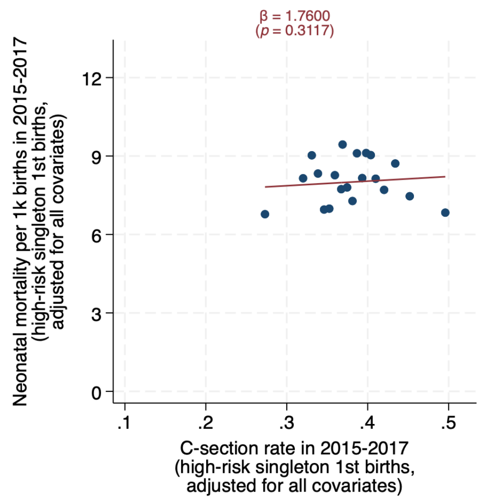
Figure 7: Correlation of Adjusted County C-Section Rates and Neonatal Mortality

High-Risk Singleton 1st Births

(a) 1989-1991

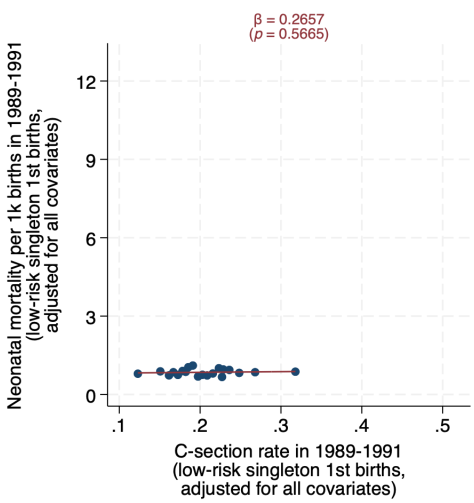


(b) 2015-2017

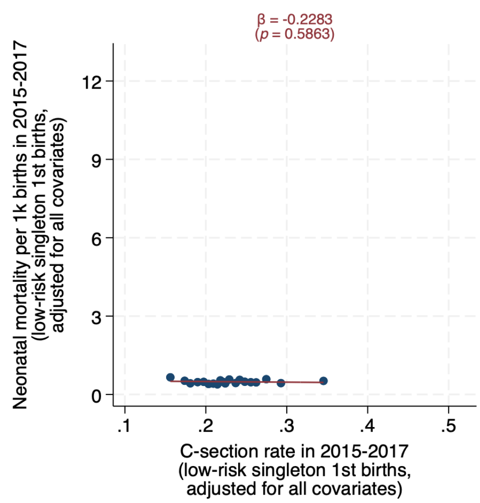


Low-Risk Singleton 1st Births

(c) 1989-1991



(d) 2015-2017

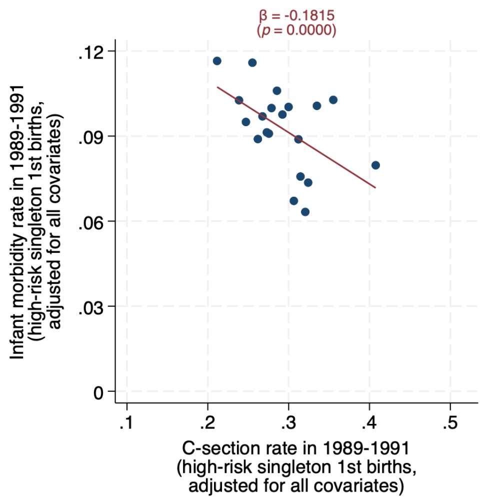


Notes: This figure shows binscatter plots of C-section rates with neonatal mortality rates. Linear fit and p-value are based on the underlying counties (prior to binning). C-section and neonatal mortality rates (adjusted for all covariates) are averaged over the three-year period, weighted by the number of relevant births in each year. Binscatter and linear fit are weighted by the number of relevant births in the county over the three years.

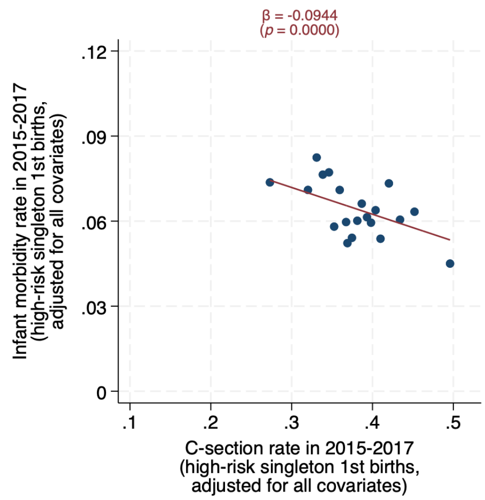
Figure 8: Correlation of Adjusted County C-Section Rates and Infant Morbidity

High-Risk Singleton 1st Births

(a) 1989-1991

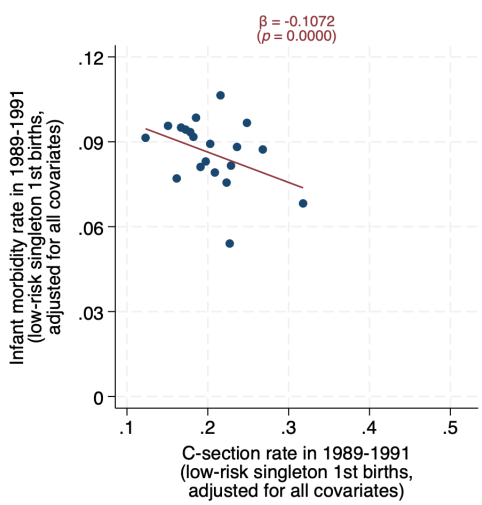


(b) 2015-2017

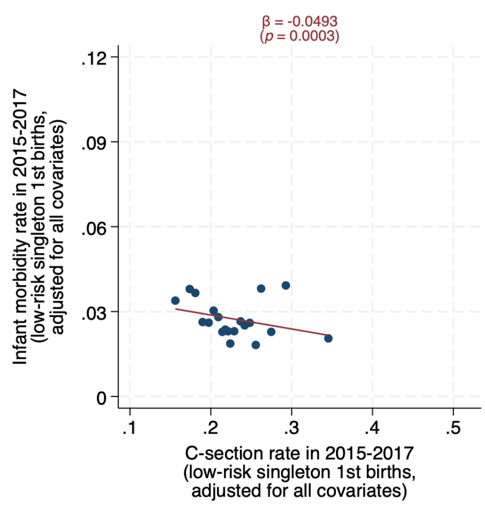


Low-Risk Singleton 1st Births

(c) 1989-1991



(d) 2015-2017

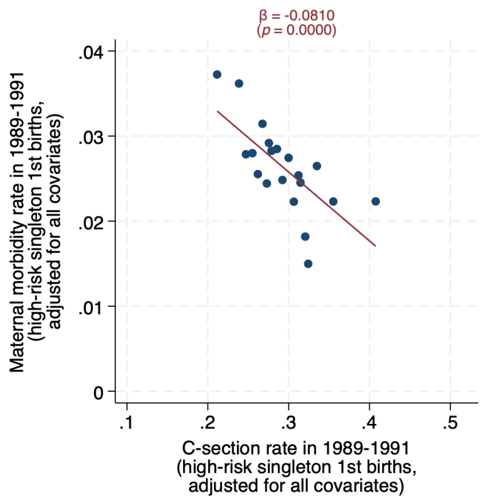


Notes: This figure shows binscatter plots of C-section rates for singleton first births with infant morbidity. Linear fit and p-value are based on the underlying counties (prior to binning). Infant morbidity is the presence of any of the following: meconium, injury, seizure, ventilation. C-section and morbidity rates (adjusted for all covariates) are averaged over the three-year period, weighted by the number of singleton first births in each year. Binscatter and linear fit are weighted by the number of singleton first births in the county over the three years.

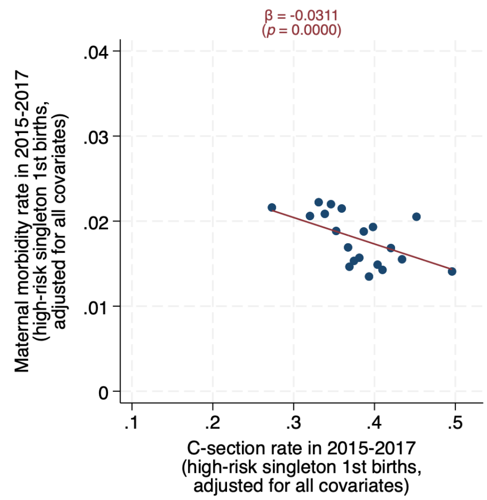
Figure 9: Correlation of Adjusted County C-Section Rates and Maternal Morbidity

High-Risk Singleton 1st Births

(a) 1989-1991

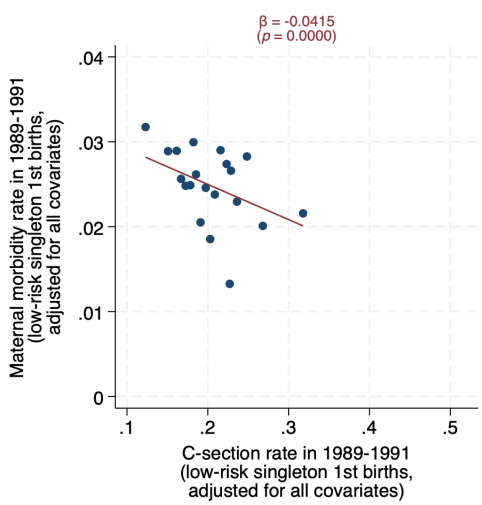


(b) 2015-2017

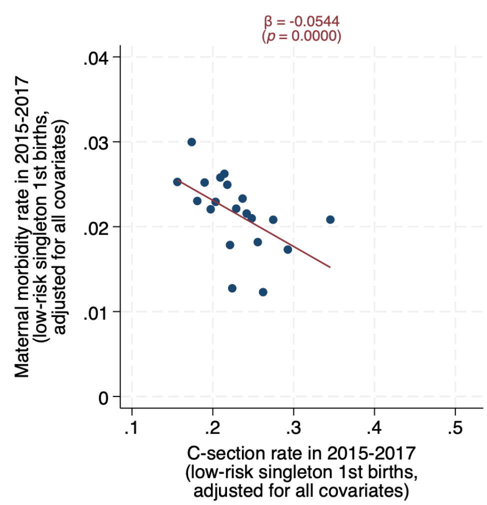


Low-Risk Singleton 1st Births

(c) 1989-1991



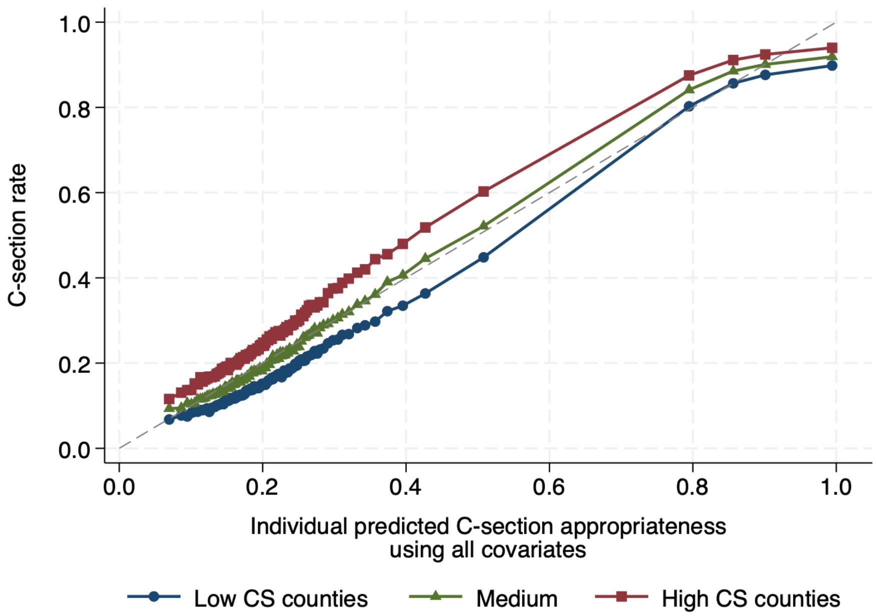
(d) 2015-2017



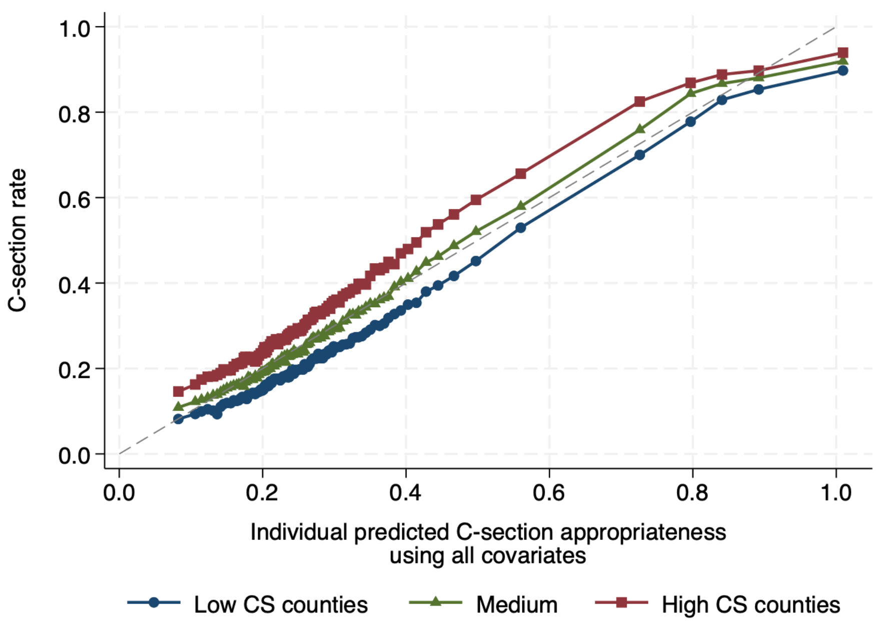
Notes: This figure shows binscatter plots of C-section rates for singleton first births with maternal morbidity. Linear fit and p-value are based on the underlying counties (prior to binning). Maternal morbidity is the presence of any of the following: febrile, excessive bleeding, seizure, transfusion, perineal laceration, ruptured uterus, unplanned hysterectomy, admission to ICU. C-section and morbidity rates (adjusted for all covariates) are averaged over the three-year period, weighted by the number of high-risk singleton first births in each year. Binscatter and linear fit are weighted by the number of singleton first births in the county over the three years.

Figure 10: C-Section Rates by Predicted C-Section Appropriateness
and County C-Section Rate

(a) 1989-1991



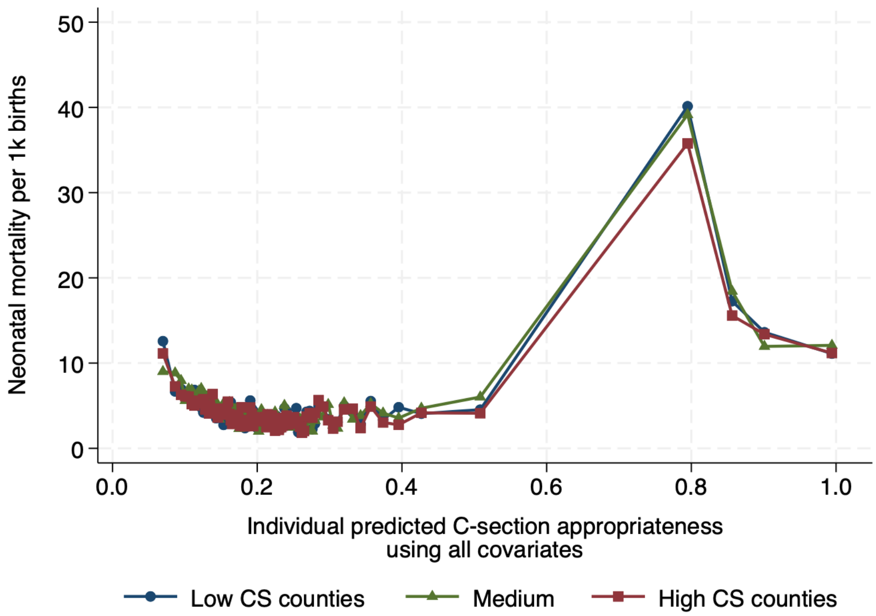
(b) 2015-2017



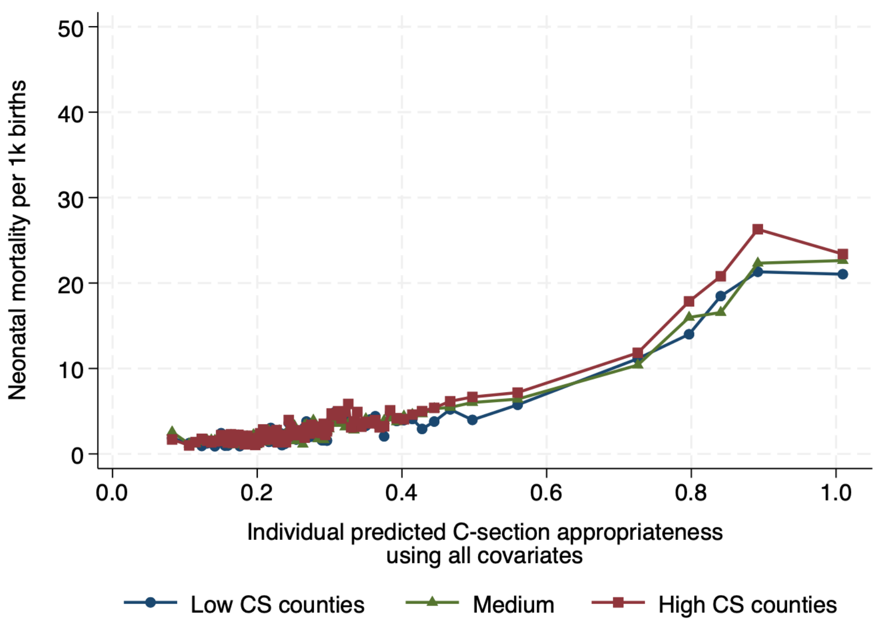
Notes: This figure shows (raw) C-section rates for singleton first births in each percentile of predicted C-section appropriateness. This is shown separately for births that take place in counties with low, medium, and high adjusted C-section rates (terciles weighted by number of singleton first births). Predicted C-section appropriateness is derived from a regression based on all covariates and county fixed effects, but the prediction excludes the county fixed effects. Note each marker in the figure represents a percentile (i.e., there are 100 markers for each curve).

Figure 11: Neonatal Mortality by Predicted C-Section Appropriateness
and County C-Section Rate

(a) 1989-1991



(b) 2015-2017

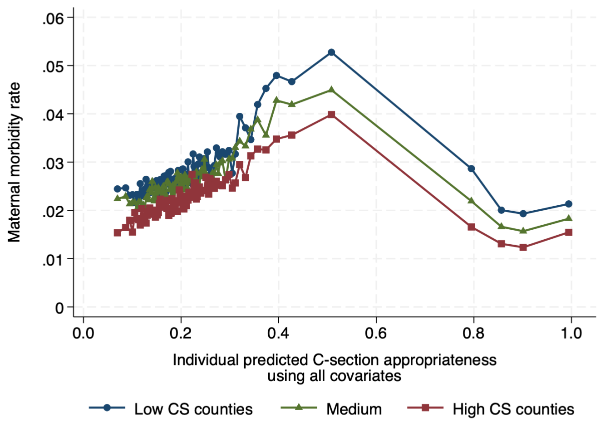


Notes: This figure shows (raw) neonatal mortality rates for singleton first births in each percentile of predicted C-section appropriateness. This is shown separately for births that take place in counties with low, medium, and high adjusted C-section rates (terciles weighted by number of singleton first births). Predicted C-section appropriateness is derived from a regression based on all covariates and county fixed effects, but the prediction excludes the county fixed effects. Note each marker in the figure represents a percentile (i.e., there are 100 markers for each curve).

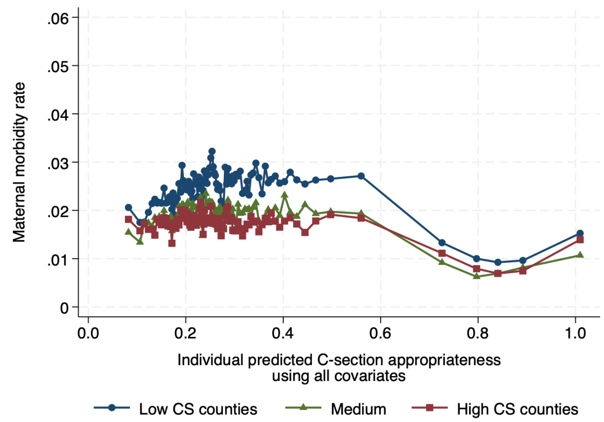
Figure 12: Maternal and Infant Morbidity by Predicted C-Section Appropriateness and County C-Section Rate

Maternal Morbidity

(a) 1989-1991

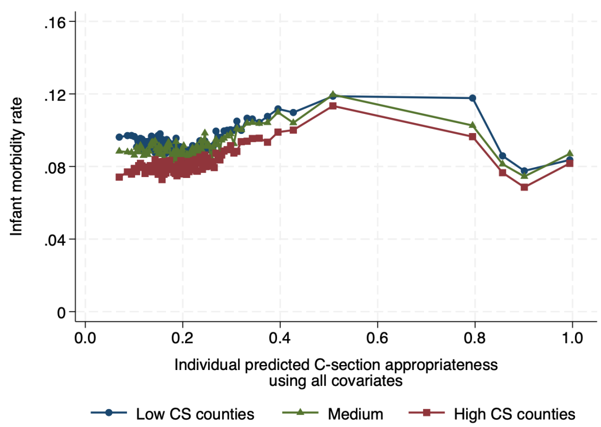


(b) 2015-2017

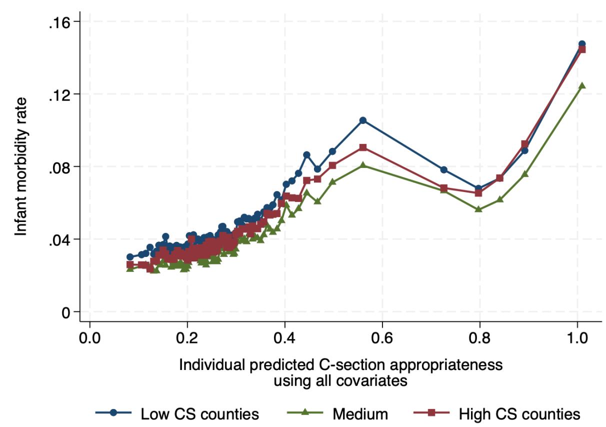


Infant Morbidity

(c) 1989-1991

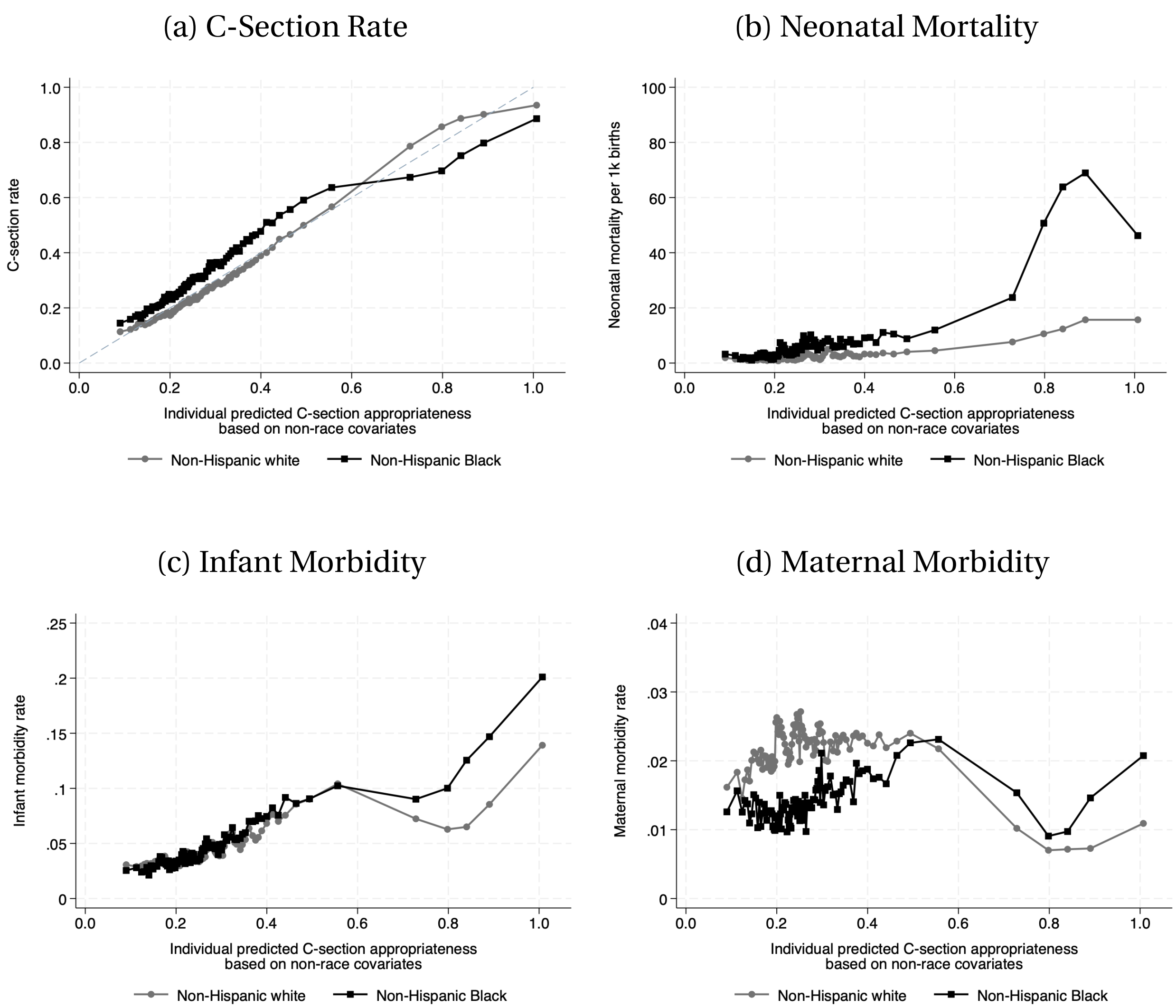


(d) 2015-2017



Notes: This figure shows (raw) maternal and infant morbidity rates for singleton first births in each percentile of predicted C-section appropriateness. This is shown separately for births that take place in counties with low, medium, and high adjusted C-section rates (terciles weighted by number of singleton first births). Predicted C-section appropriateness is derived from a regression based on all covariates and county fixed effects, but the prediction excludes the county fixed effects. Maternal morbidity is the presence of any of the following: febrile, excessive bleeding, seizure, transfusion, perineal laceration, ruptured uterus, unplanned hysterectomy, admission to ICU. Infant morbidity is the presence of any of the following: meconium, injury, seizure, ventilation. Note each marker in the figure represents a percentile (i.e., there are 100 markers for each curve).

Figure 13: C-Section Rates, Neonatal Mortality, and Maternal and Infant Morbidity
by Predicted C-Section Appropriateness and Race, 2015-2017
(Based on Non-Race Covariates)



Notes: This figure shows (raw) C-section, neonatal mortality, and maternal and infant morbidity rates for singleton first births in each percentile of predicted C-section appropriateness. This is shown separately for births with non-Hispanic white mothers and non-Hispanic Black mothers. Predicted C-section appropriateness is derived from a regression based on non-race covariates and county fixed effects, but the prediction excludes the county fixed effects. Maternal morbidity is the presence of any of the following: febrile, excessive bleeding, seizure, transfusion, perineal laceration, ruptured uterus, unplanned hysterectomy, admission to ICU. Infant morbidity is the presence of any of the following: meconium, injury, seizure, ventilation. Note each marker in the figure represents a percentile (i.e., there are 100 markers for each curve).

APPENDIX

A Selection of Covariates

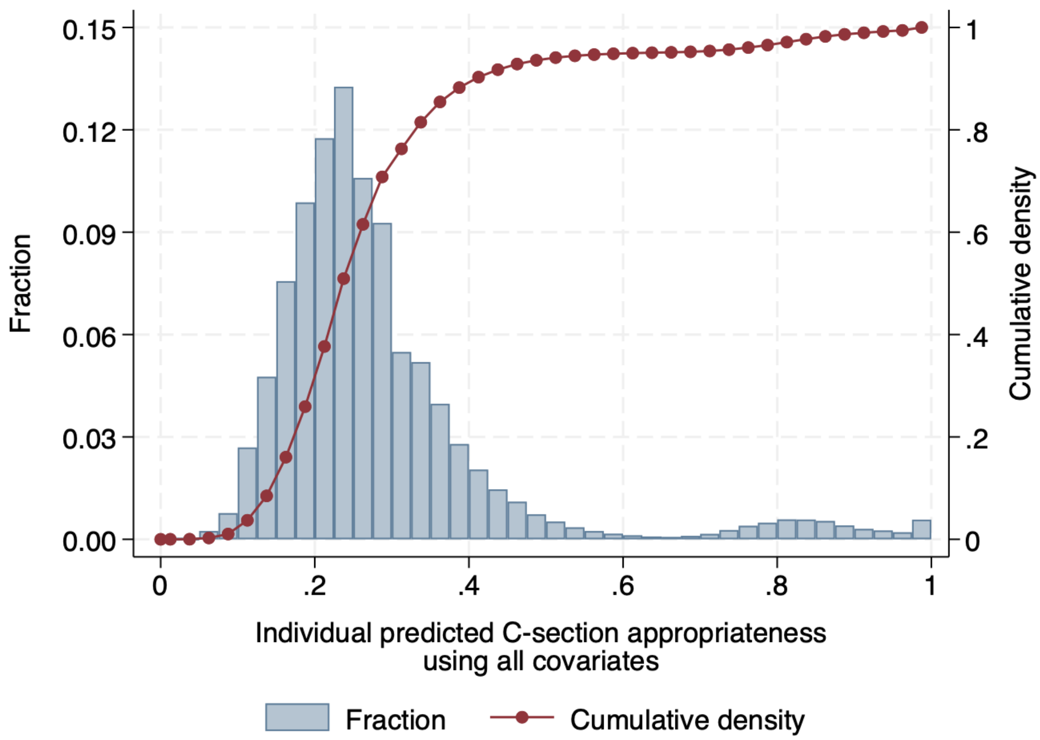
Table A.1: Comparison of Covariates with Select Literature

	This paper			
	High-risk (vs. low-risk) births	Predicted C-section appropriateness	Card, Fenizia, and Silver (2023)	Currie and MacLeod (2017)
Maternal age	<18 or >35	5-year bins	<18 or >35	5-year bins
Gestational age	< 37 weeks	< 37 weeks	< 37 weeks	–
Prenatal visits	≥ 19	≥ 19	> 20	–
Growth restrictions	✓	✓	✓	–
Breech	✓	✓	✓	✓
Eclampsia	✓	✓	✓	–
Pre-eclampsia	✓	✓	✓	–
Diabetes	✓	✓	–	–
BMI (avail. $\geq$ 2009)	–	*See Table 1*	90$^{\text{th}}$ pctile	–
Placenta previa (avail. $\leq$ 2006)	–	–	–	✓
Abruptio placenta (avail. $\leq$ 2006)	–	–	–	✓
Cord prolapse (avail. $\leq$ 2006)	–	–	–	✓
Multiple birth	N/A (singleton 1$^{\text{st}}$ births)		✓	✓
Birth order	N/A (singleton 1$^{\text{st}}$ births)		Non-first	✓
Previous C-section	N/A (singleton 1$^{\text{st}}$ births)		N/A	✓
Previous large infant	N/A (singleton 1$^{\text{st}}$ births)		N/A	✓
Previous preterm	N/A (singleton 1$^{\text{st}}$ births)		N/A	✓
Non-medical factors	–	*See Table 1*	–	–

Notes: This table describes the sets of covariates used to define high- vs. low-risk births; to adjust county rates of C-section, neonatal mortality, and maternal & infant morbidity; and to predict C-section appropriateness for individuals. See Table 1 for additional information about the use of covariates in this paper.

B Predicted C-Section Appropriateness

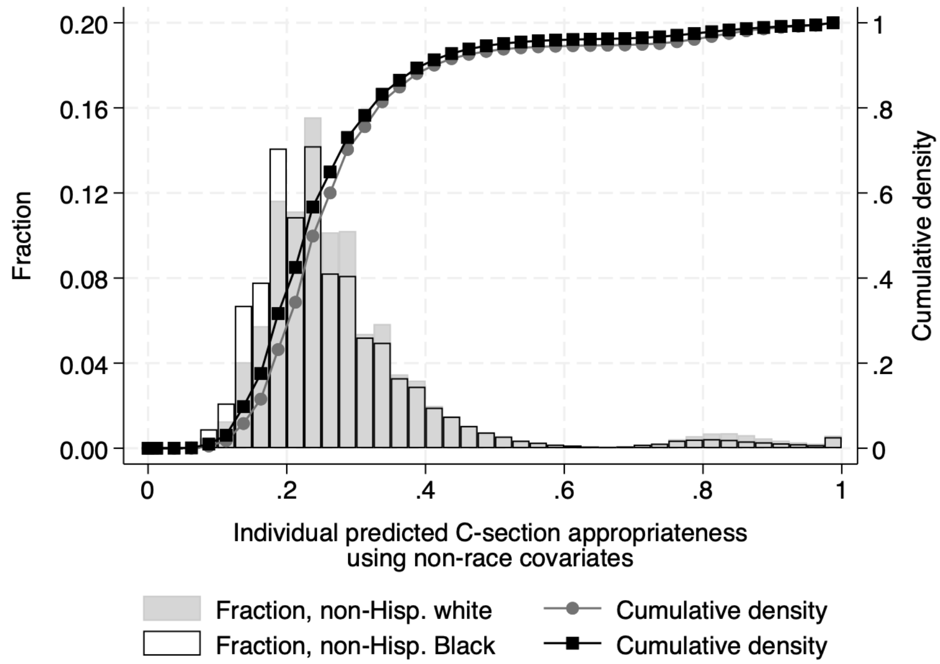
Figure B.1: Individual Predicted C-Section Appropriateness 2015-2017



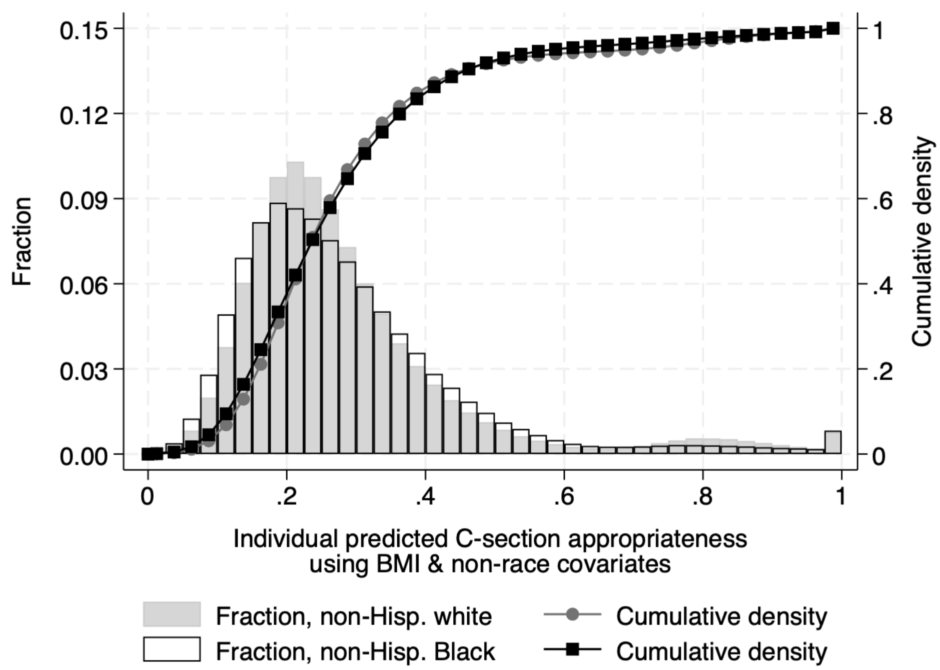
Notes: This figure shows the distribution of the individual predicted C-section appropriateness for singleton first births using all covariates (after controlling for county fixed effects).

Figure B.2: Individual Predicted C-Section Appropriateness by Race 2015-2017

(a) Based on Non-Race Covariates



(b) Based on Non-Race Covariates plus BMI



Notes: This figure shows the distribution of the individual predicted C-section appropriateness for singleton first births using non-race covariates and non-race covariates plus BMI (after controlling for county fixed effects).

C Using Medical Covariates Only

Table C.1: Distribution of Births Across Low- and High-Risk
by Predicted C-Section Appropriateness
(Adjusted for Medical Covariates Only)

(a) 1989-1991

	Predicted C-section appropriateness (using medical covariates)		
	≤ 0.6	*> 0.6*	*Overall*
Low-risk births	63.32 %	0.00 %	60.74 %
High-risk births	33.28 %	100.00 %	35.99 %
Unknown risk births	3.40 %	0.00 %	3.27 %
	100%	100%	100%

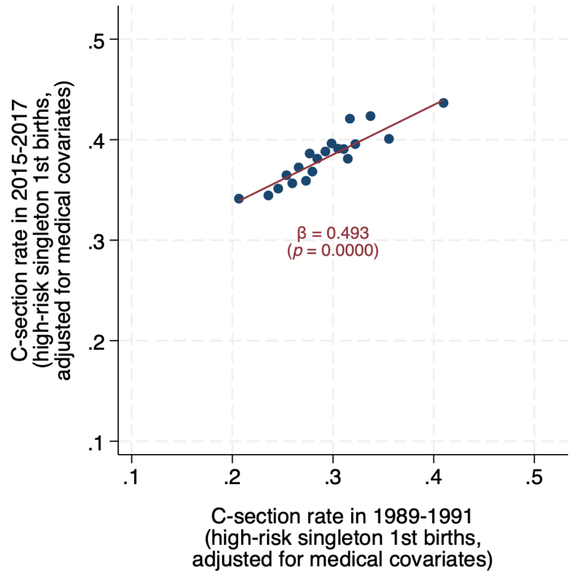
(b) 2015-2017

	Predicted C-section appropriateness (using medical covariates)		
	≤ 0.6	*> 0.6*	*Overall*
Low-risk births	63.64 %	0.00 %	60.41 %
High-risk births	34.05 %	100.00 %	37.40 %
Unknown risk births	2.31 %	0.00 %	2.19 %
	100%	100%	100%

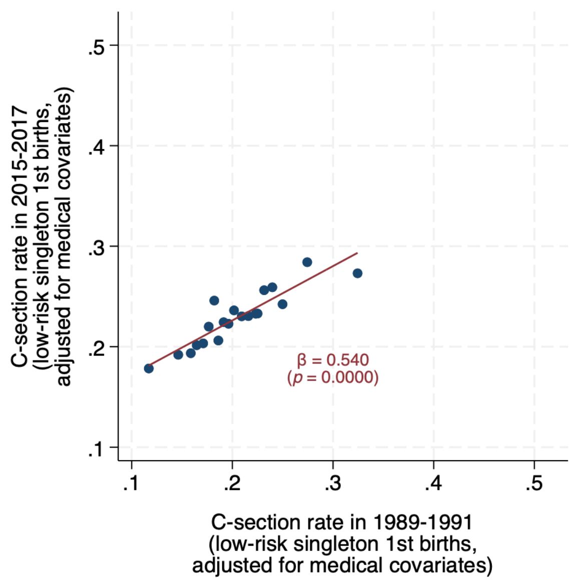
Notes: This table shows the relationship between two different approaches for assessing risk of C-section: (1) the categorization of births as low- or high-risk, a binary assignment based on medical factors only, and (2) the predicted C-section appropriateness, a continuum of risk estimated using the same medical factors. See Table 1 for the specific covariates used in each model. Only singleton first births are represented. A small portion of singleton first births have no observed high-risk characteristics but are missing data, and thus are not classified as either low- or high-risk.

Figure C.1: Persistence in County C-Section Rates Over Time
(Adjusted for Medical Covariates Only)

(a) High-Risk Singleton 1st Births



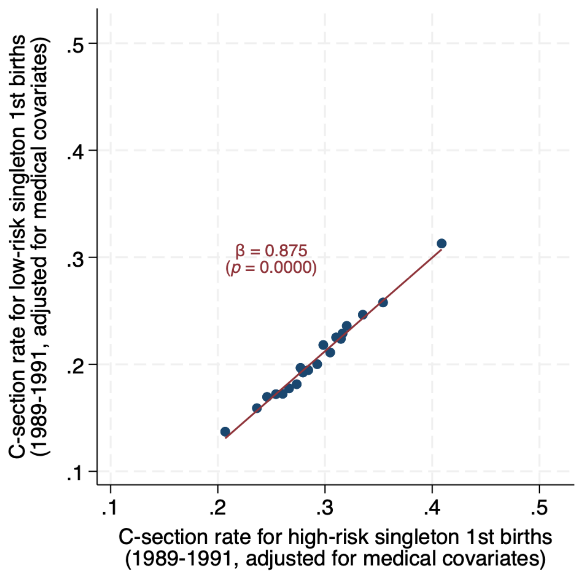
(b) Low-Risk Singleton 1st Births



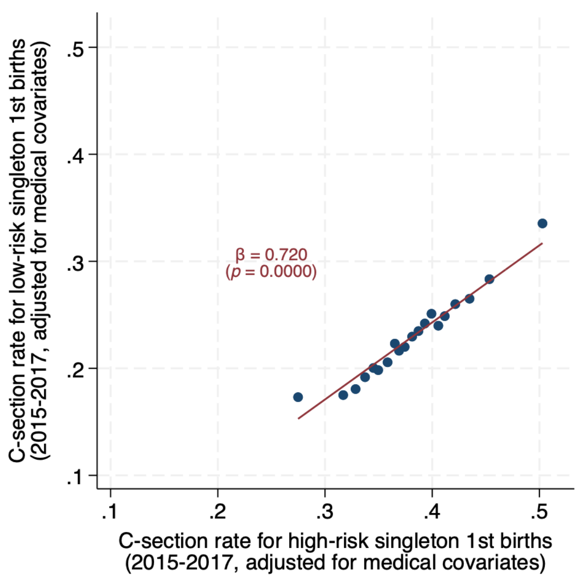
Notes: This figure shows binscatter plots of C-section rates across time periods. Linear fit and p-value are based on the underlying counties (prior to binning). C-section rates (adjusted for medical covariates) are averaged over the three-year period, weighted by the number of relevant births in each year. Binscatter and linear fit are weighted by the number of relevant births in the county over all six years.

Figure C.2: Correlation in County C-Section Rates Across Risk Type
(Adjusted for Medical Covariates Only)

(a) 1989-1991



(b) 2015-2017

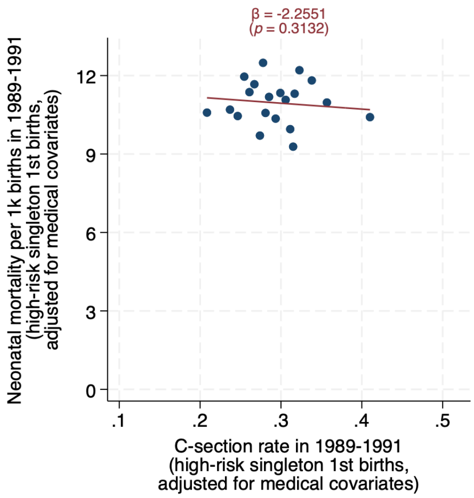


Notes: This figure shows binscatter plots of C-section rates across risk types. Linear fit and p-value are based on the underlying counties (prior to binning). C-section rates (adjusted for medical covariates) are averaged over the three-year period, weighted by the number of relevant births in each year. Binscatter and linear fit are weighted by the number of singleton first births in the county over the three years.

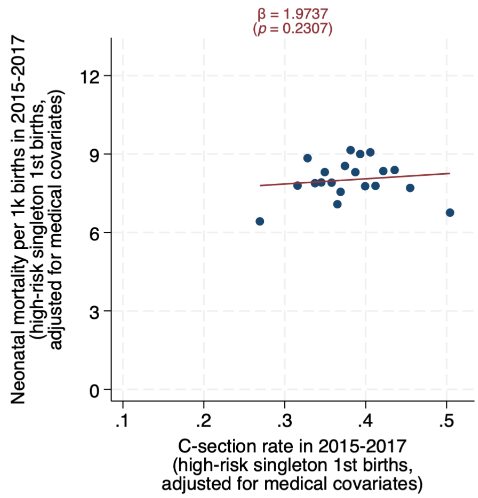
Figure C.3: Correlation of County C-Section Rates and Neonatal Mortality
(Adjusted for Medical Covariates Only)

High-Risk Singleton 1st Births

(a) 1989-1991

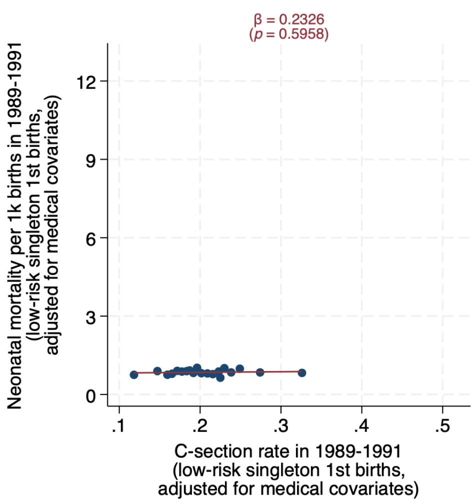


(b) 2015-2017

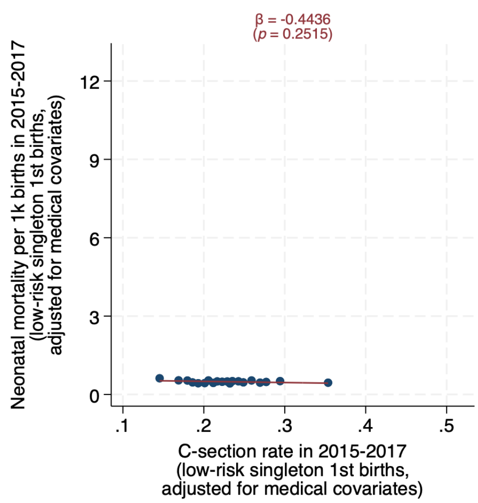


Low-Risk Singleton 1st Births

(c) 1989-1991



(d) 2015-2017

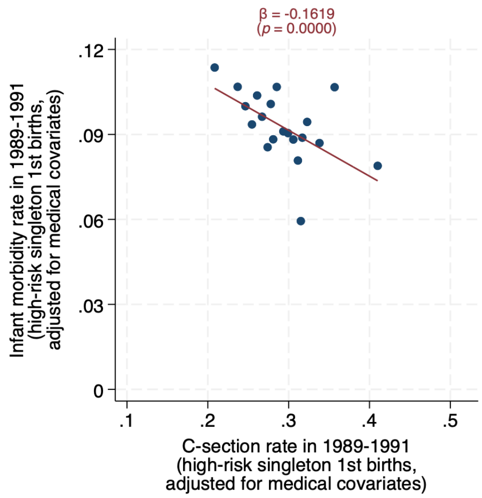


Notes: This figure shows binscatter plots of C-section rates with neonatal mortality rates. Linear fit and p-value are based on the underlying counties (prior to binning). C-section and neonatal mortality rates (adjusted for medical covariates) are averaged over the three-year period, weighted by the number of relevant births in each year. Binscatter and linear fit are weighted by the number of relevant births in the county over the three years.

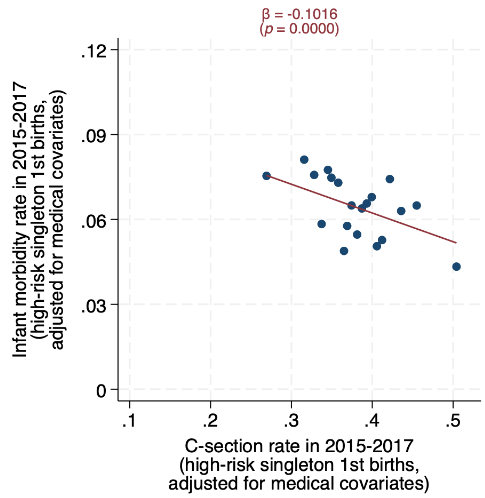
Figure C.4: Correlation of County C-Section Rates and Infant Morbidity (Adjusted for Medical Covariates Only)

High-Risk Singleton 1st Births

(a) 1989-1991

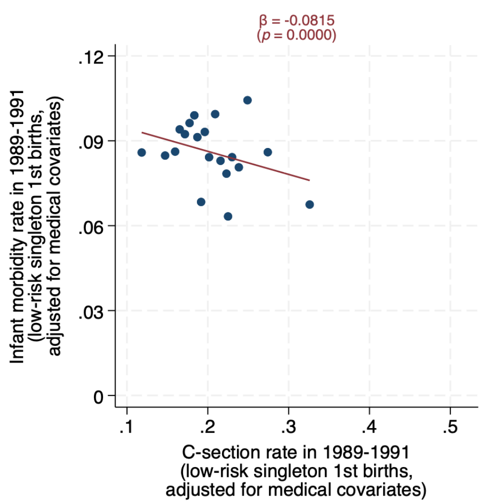


(b) 2015-2017

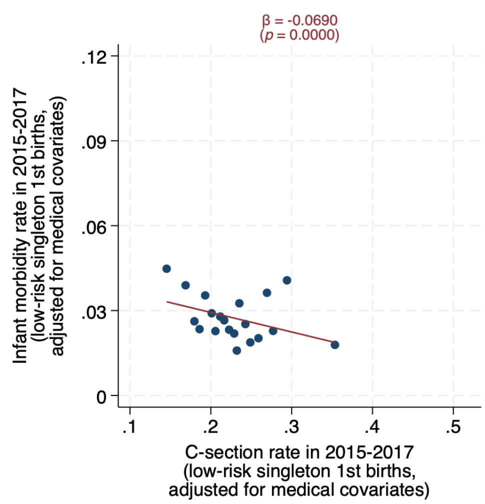


Low-Risk Singleton 1st Births

(c) 1989-1991



(d) 2015-2017

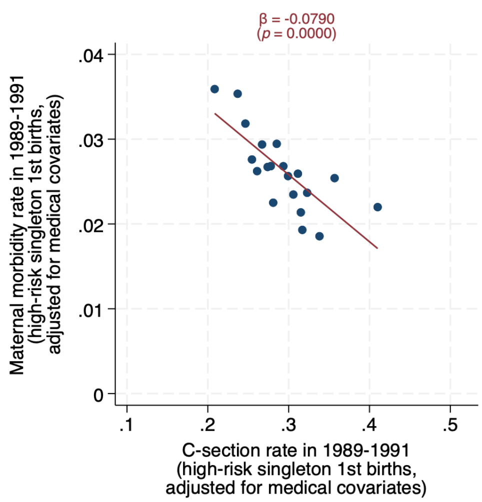


Notes: This figure shows binscatter plots of C-section rates for low-risk singleton first births with maternal or infant morbidity. Linear fit and p-value are based on the underlying counties (prior to binning). Infant morbidity is the presence of any of the following: meconium, injury, seizure, ventilation. C-section and morbidity rates (adjusted for all covariates) are averaged over the three-year period, weighted by the number of low-risk singleton first births in each year. Binscatter and linear fit are weighted by the number of low-risk singleton first births in the county over the three years.

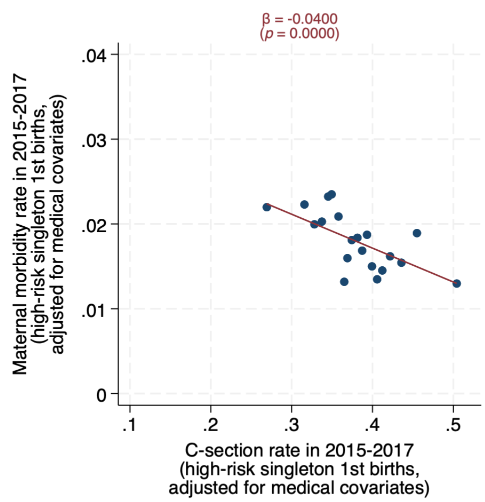
Figure C.5: Correlation of County C-Section Rates and Maternal Morbidity (Adjusted for Medical Covariates Only)

High-Risk Singleton 1st Births

(a) 1989-1991

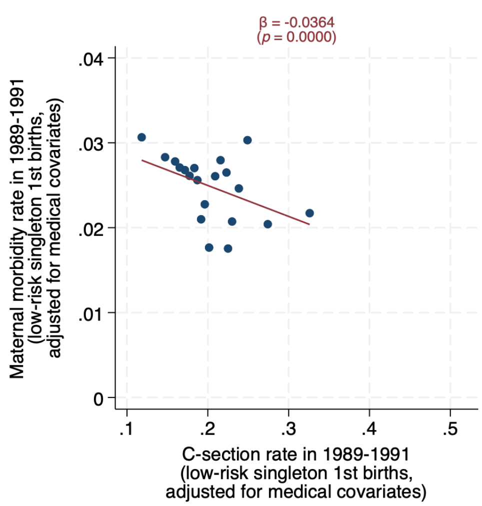


(b) 2015-2017

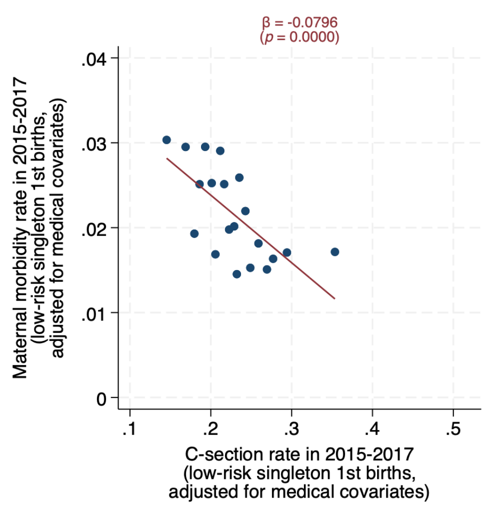


Low-Risk Singleton 1st Births

(c) 1989-1991



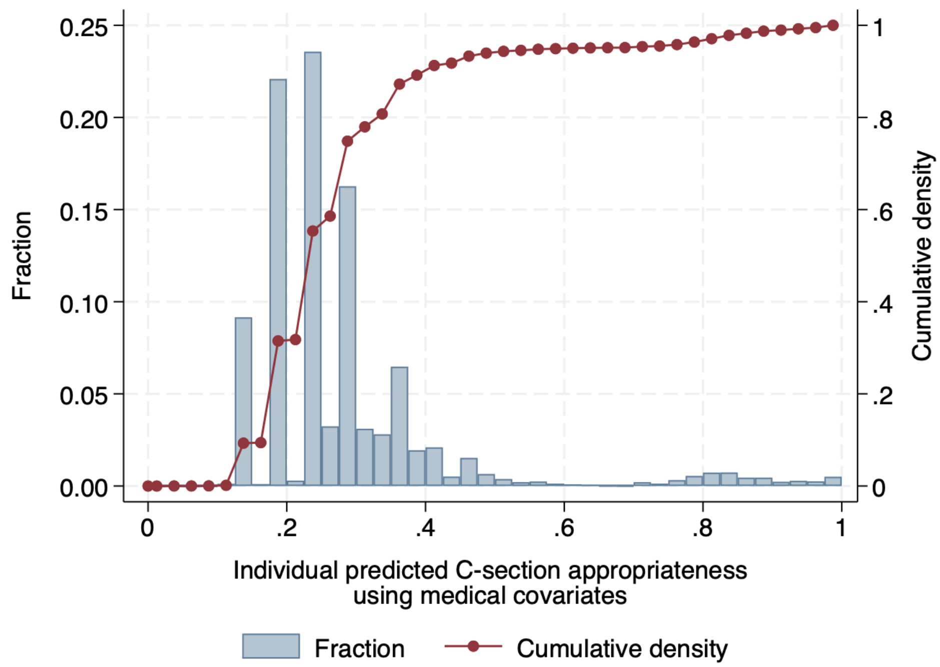
(d) 2015-2017



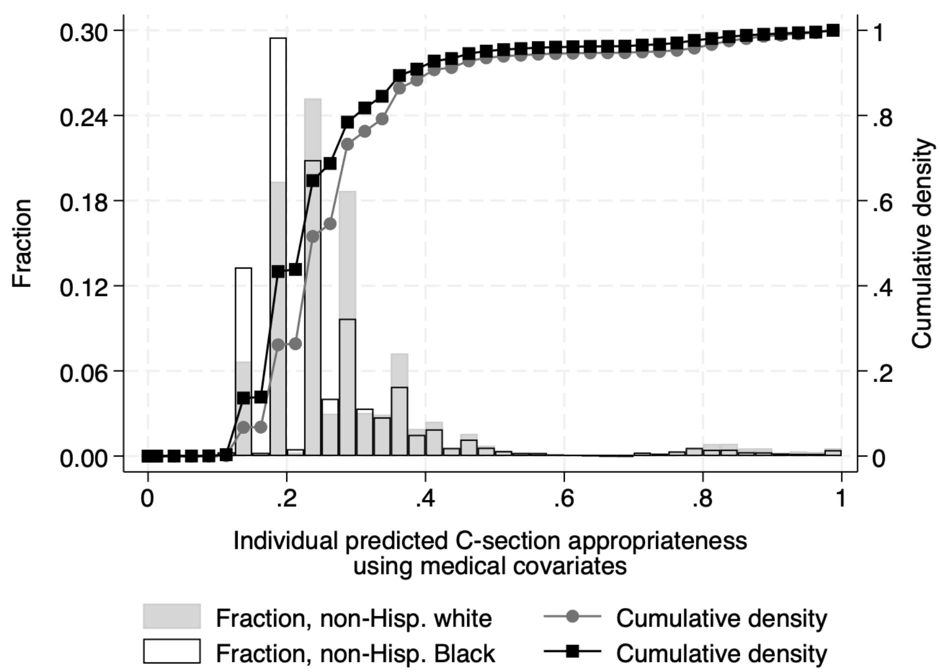
Notes: This figure shows binscatter plots of C-section rates for high-risk singleton first births with maternal or infant morbidity. Linear fit and p-value are based on the underlying counties (prior to binning). Maternal morbidity is the presence of any of the following: febrile, excessive bleeding, seizure, transfusion, perineal laceration, ruptured uterus, unplanned hysterectomy, admission to ICU. C-section and morbidity rates (adjusted for medical covariates) are averaged over the three-year period, weighted by the number of high-risk singleton first births in each year. Binscatter and linear fit are weighted by the number of high-risk singleton first births in the county over the three years.

Figure C.6: Individual Predicted C-Section Appropriateness Overall and by Race
2015-2017
(Based on Medical Covariates Only)

(a) Overall



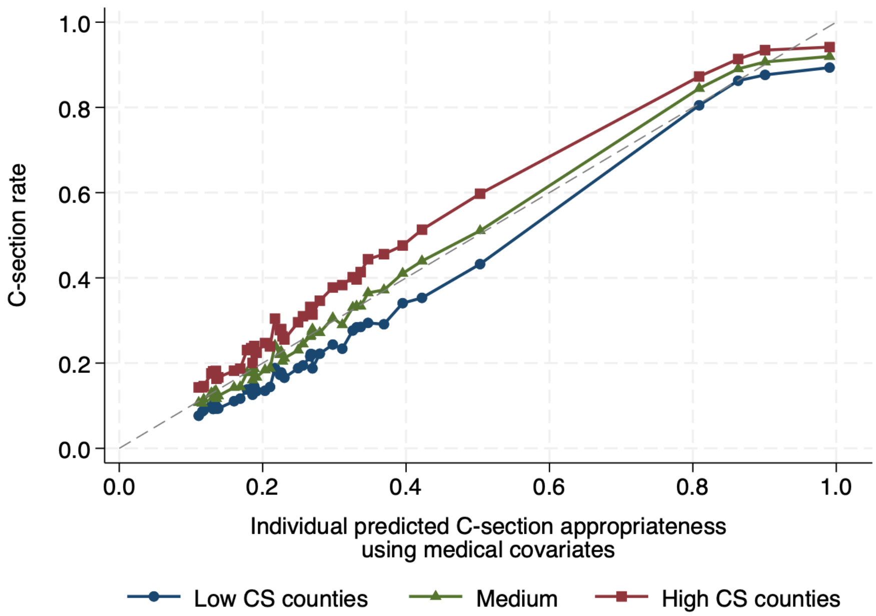
(b) By Race



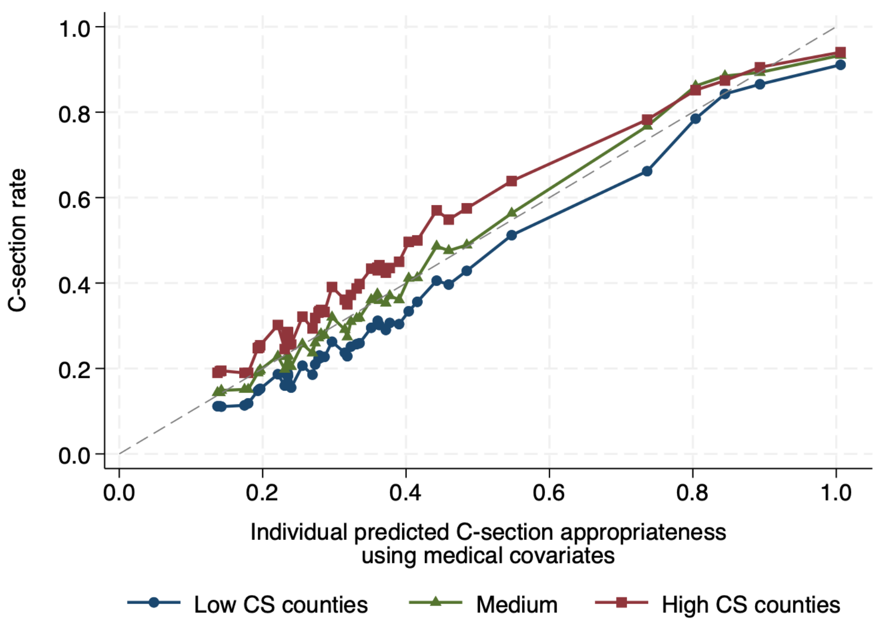
Notes: These figures show the distribution of the individual predicted C-section appropriateness for singleton first births using medical covariates (after controlling for county fixed effects).

Figure C.7: C-Section Rates by Predicted C-Section Appropriateness
and County C-Section Rate
(Based on Medical Covariates Only)

(a) 1989-1991



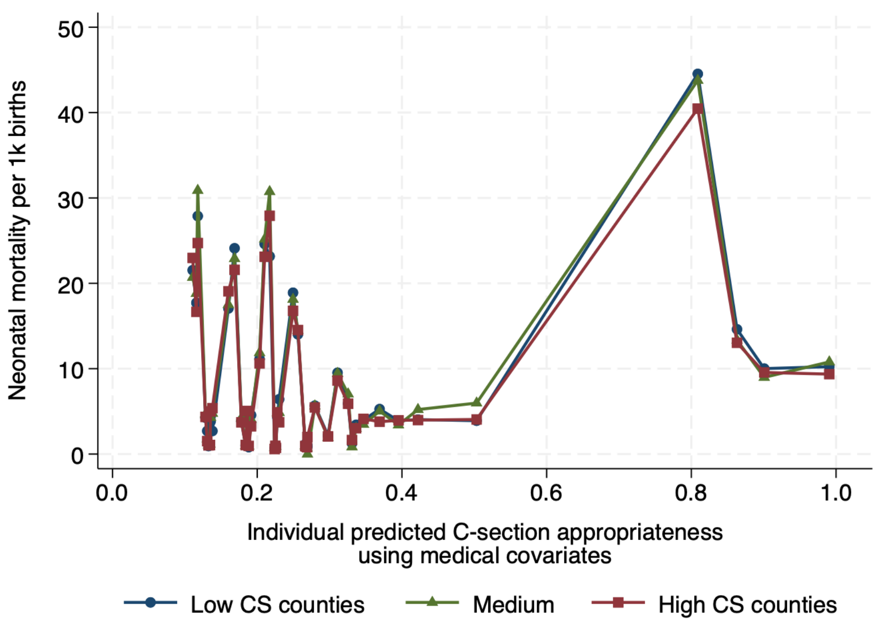
(b) 2015-2017



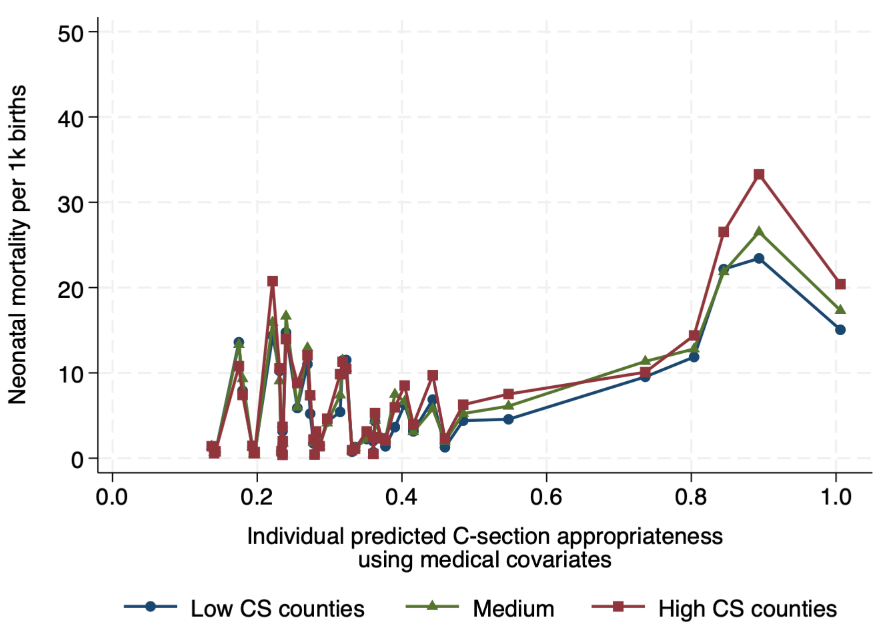
Notes: This figure shows (raw) C-section rates for singleton first births in each percentile of predicted C-section appropriateness. This is shown separately for births that take place in counties with low, medium, and high adjusted C-section rates (terciles weighted by number of singleton first births). Predicted C-section appropriateness is derived from a regression based on medical covariates and county fixed effects, but the prediction excludes the county fixed effects.

Figure C.8: Neonatal Mortality by Predicted C-Section Appropriateness
and County C-Section Rate
(Based on Medical Covariates Only)

(a) 1989-1991



(b) 2015-2017

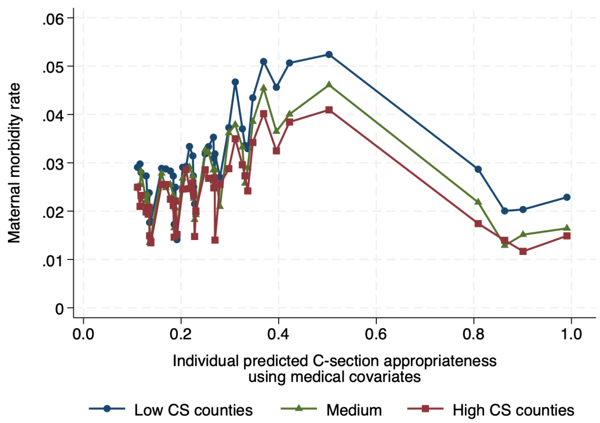


Notes: This figure shows (raw) neonatal mortality rates for singleton first births in each percentile of predicted C-section appropriateness. This is shown separately for births that take place in counties with low, medium, and high adjusted C-section rates (terciles weighted by number of singleton first births). Predicted C-section appropriateness is derived from a regression based on medical covariates and county fixed effects, but the prediction excludes the county fixed effects.

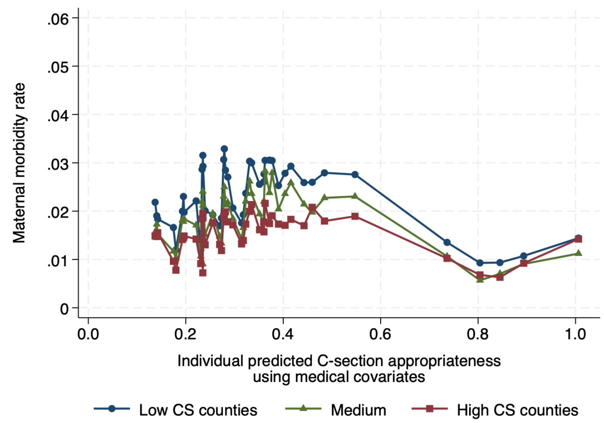
Figure C.9: Maternal and Infant Morbidity by Predicted C-Section Appropriateness and County C-Section Rate
(Based on Medical Covariates Only)

Maternal Morbidity

(a) 1989-1991

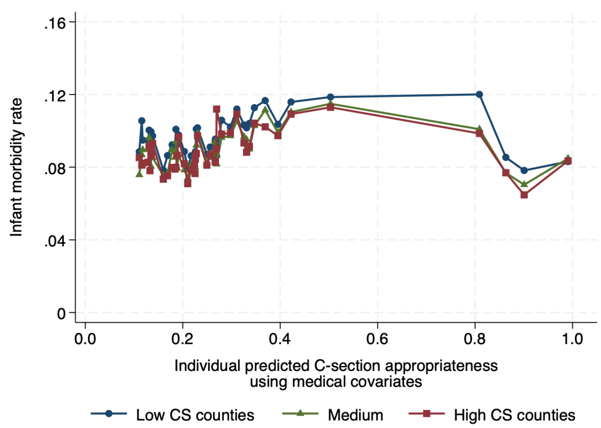


(b) 2015-2017

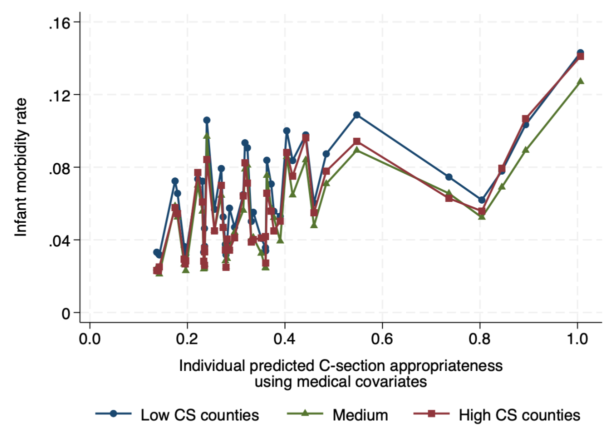


Infant Morbidity

(c) 1991-1993

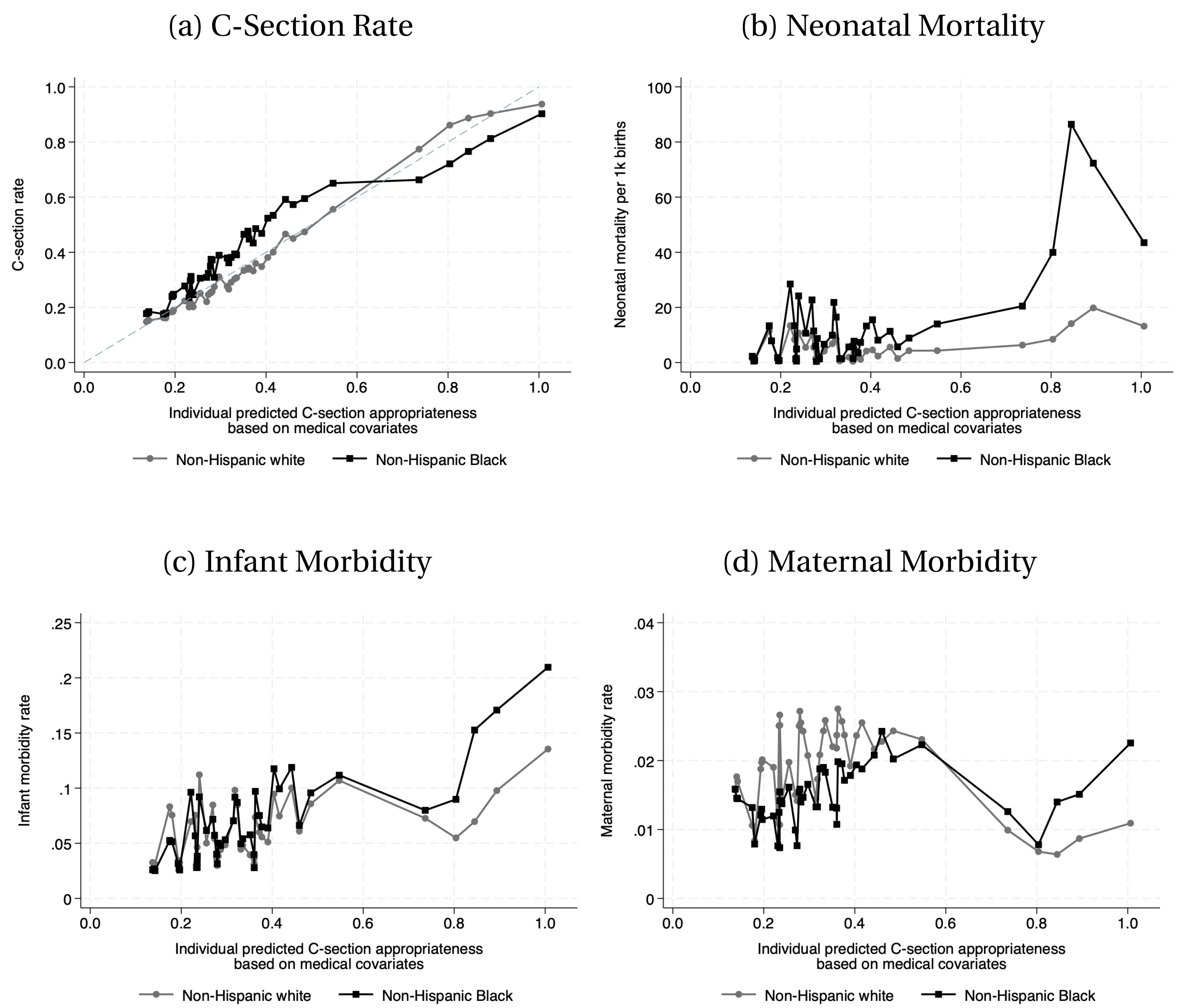


(d) 2015-2017



Notes: This figure shows (raw) maternal and infant morbidity rates for singleton first births in each percentile of predicted C-section appropriateness. This is shown separately for births that take place in counties with low, medium, and high adjusted C-section rates (terciles weighted by number of singleton first births). Predicted C-section appropriateness is derived from a regression based on medical covariates and county fixed effects, but the prediction excludes the county fixed effects. Maternal morbidity is the presence of any of the following: febrile, excessive bleeding, seizure, transfusion, perineal laceration, ruptured uterus, unplanned hysterectomy, admission to ICU. Infant morbidity is the presence of any of the following: meconium, injury, seizure, ventilation.

Figure C.10: C-Section Rates, Neonatal Mortality, and Maternal and Infant Morbidity by Predicted C-Section Appropriateness and Race, 2015-2017 (Based on Medical Covariates Only)

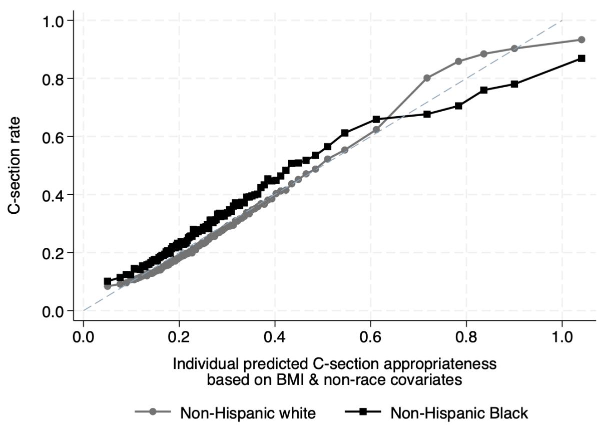


Notes: This figure shows (raw) C-section, neonatal mortality, and maternal and infant morbidity rates for singleton first births in each percentile of predicted C-section appropriateness. This is shown separately for births with non-Hispanic white mothers and non-Hispanic Black mothers. Predicted C-section appropriateness is derived from a regression based on medical covariates and county fixed effects, but the prediction excludes the county fixed effects. Maternal morbidity is the presence of any of the following: febrile, excessive bleeding, seizure, transfusion, perineal laceration, ruptured uterus, unplanned hysterectomy, admission to ICU. Infant morbidity is the presence of any of the following: meconium, injury, seizure, and ventilation.

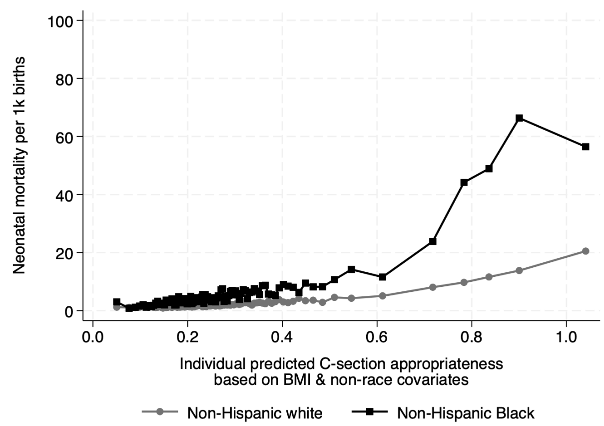
D Using Body Mass Index

Figure D.1: C-Section Rates, Neonatal Mortality, and Maternal and Infant Morbidity by Predicted C-Section Appropriateness and Race, 2015-2017
(Based on Non-Race plus BMI Covariates)

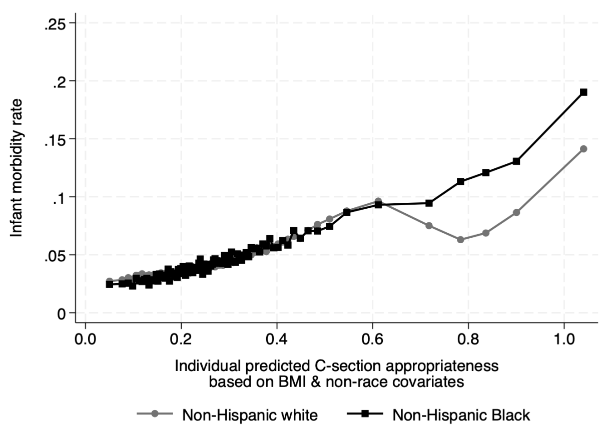
(a) C-Section Rate



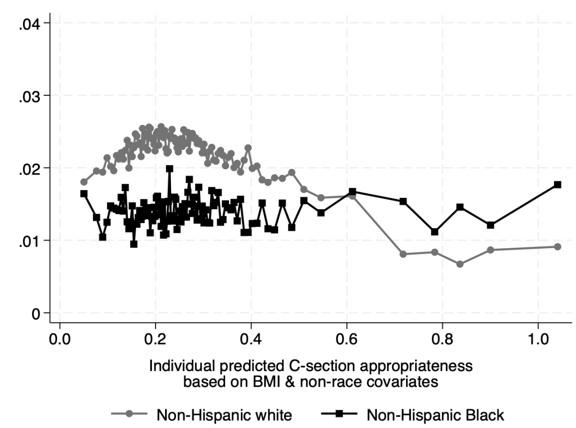
(b) Neonatal Mortality



(c) Infant Morbidity



(d) Maternal Morbidity



Notes: This figure shows (raw) C-section, neonatal mortality, and maternal and infant morbidity rates for singleton first births in each percentile of predicted C-section appropriateness. This is shown separately for births with non-Hispanic white mothers and non-Hispanic Black mothers. Predicted C-section appropriateness is derived from a regression based on non-race plus BMI covariates and county fixed effects, but the prediction excludes the county fixed effects. Maternal morbidity is the presence of any of the following: febrile, excessive bleeding, seizure, transfusion, perineal laceration, ruptured uterus, unplanned hysterectomy, admission to ICU. Infant morbidity is the presence of any of the following: meconium, injury, seizure, ventilation. Note each marker in the figure represents a percentile (i.e., there are 100 markers for each curve).