Health Care Centralization: The Health Impacts of Obstetric Unit Closures in the US *

Stefanie Fischer
Monash University and IZA
stefanie.fischer@monash.edu

Heather Royer
University of California, Santa Barbara, NBER, IZA
royer@econ.ucsb.edu

Corey White
Monash University and IZA
corey.white@monash.edu

April 11, 2022

Abstract

Over the last few decades, health care services in the United States have become more centralized. We study how the loss of hospital-based obstetric units in over 400 rural counties affect maternal and infant health via a difference-in-differences design. We find that closures lead to significant changes in obstetric practices such as inductions and C-sections. In contrast to concerns voiced in the public discourse, the effects on a range of maternal and infant health outcomes are negligible or slightly beneficial. While women travel farther to receive care, closures induce women to receive higher quality care.

JEL Codes: J13, I18, J08, J18, I38
Keywords: obstetrics, closures, infant health, maternal health

*This paper has benefited from helpful conversations with Mark Anderson, Susan Athey, Janet Currie, Joe Doyle, Gautam Gowrisankaran, Ben Handel, Peter Hull, David Molitor, Dan Rees, and Maya Rossin-Slater,. Thanks to Joe Doyle and John Graves for sharing their cleaned hospital quality data. We greatly appreciate the wonderful research assistant work of Anna Jaskiewicz and Amira Garewal. We thank participants at the following conferences and seminars for providing useful feedback: Hawaii Applied Micro One-Day Conference (2019), Society of Labor Economists Annual Meeting (2020), American Economic Association Annual Meetings (2021), University of Melbourne (2022), and Monash University (2022).

1 Introduction

In the past three decades, over 400 counties have lost their sole hospital-based obstetric (OB) unit.[1] Today only about half of all US counties have a hospital-based OB unit within their borders. These closures are part of a trend of the regionalization of perinatal services, beginning in the 1970s, whereby advanced neonatal technologies became more centralized. These closures have disproportionately impacted vulnerable communities with high rates of Medicaid usage, elevated rates of poverty, and a larger fraction of black female residents (Hung et al., 2017).

The loss of OB services, particularly acute for the 60 million people living in rural communities in the United States, has garnered considerable public policy attention —characterized as the "Rural Maternity Care Crisis" (Commonwealth, 2019). The Centers for Medicare and Medicaid Services (CMS) has taken action through its creation of the Rural Health Council in 2016 with the goal of ensuring access to high-quality health care to rural Americans.[2]

The most direct consequence of these closures, and the one of focal interest, is the reduction in the proximity of health care services.[3] When an OB unit closes, many pregnant women must travel farther to receive care —both prior to delivery and at the time of delivery. For counties that lost their only OB unit, the distance to the nearest unit increased by over 30 miles on average.[4] An increase in the distance to care may lead to higher rates of labor and delivery complications —having implications for both the mother and the newborn. However, when an OB unit closes, a pregnant woman must decide on an alternative health care provider. The new health care provider may provide better or worse services than the hospital with the closed OB unit. On net, the impact of an OB unit closure is unclear, especially if women are redirected to hospitals with higher quality care.

In this paper, we study how these OB unit closures affect maternal and infant health outcomes. Specifically, we leverage 1989-2019 within-county variation in the existence of at least one OB unit in the county via a dynamic difference-in-differences design. We appease worries about the comparability of closure and non-closure counties in two main ways. First, we present all estimates in an event-study framework and look for changes in the outcomes that coincide precisely with the timing of treatment. Second, we supplement our main estimates with propensity-weighted difference-in-differences estimates, which take into account that based on observables, some counties may be more likely to experience a closure than others.

Our empirical analysis yields several key findings. The closures induce fewer women to de-

[1] Authors' calculation.
[2] https://www.cms.gov/About-CMS/Agency-Information/OMH/equity-initiatives/rural-health/rural-maternal-health.
[3] https://www.scientificamerican.com/article/maternal-health-care-is-disappearing-in-rural-america/.
[4] Authors' calculation.

liver in their county of residence (30 percentage point decrease), reduce the number of prenatal care visits (0.17 fewer visits), and increase the probability of slightly earlier delivery due to the raised likelihood of scheduled induction (one percentage point increase). We then examine more downstream outcomes that are plausibly affected by characteristics of the birth hospital. Closures lead mothers to experience a one percentage point reduced chance of C-section, and no statistically-significant harms to several measures of maternal or infant health, including mortality. If anything, we find a small improvement in maternal health (0.1 standard deviation improvement in our maternal morbidity index).

We next investigate several possible mechanisms and conclude that reallocation to hospitals with different characteristics is likely the dominant mechanism explaining the effects on C-sections and maternal morbidity. On average, closures induce mothers to give birth in counties that have: lower risk-adjusted C-section rates (one percentage point decrease), higher quality hospitals (0.1 standard deviation increase in our hospital quality index), and more obstetric-specific resources (four percentage points more likely to have a large neonatal intensive care unit). Exploiting heterogeneity across the large number of closures, we find that the impacts on C-sections are largest for the closures that divert women to counties with the lowest C-section rates, emphasizing the importance of place-based effects in health care (Deryugina and Molitor, 2021). Similarly, the reduction in maternal morbidity is most sizable for the closures most likely to redirect mothers to counties with more obstetric resources.

We contribute to the small collection of studies of obstetric unit closures, a common phenomenon across many developed countries. Focusing on US closures over a shorter period (2004 to 2014) and a narrower set of outcomes, Kozhimannil et al. (2018) use an interrupted time series design with state fixed effects. They conclude that the closures shifted women to give birth in hospitals without obstetric units and resulted in higher rates of premature births (we find no impact on prematurity using our methods and longer time frame). In addition to the different methods, time frame, and outcomes, a critical difference with our work is our emphasis on the role of hospital attributes in understanding the effect of closures. Looking at maternity ward closures in Sweden, the working paper Avdic et al. (2020) uncover positive effects for infants but negative impacts for mothers. They postulate that hospital overcrowding is a contributing factor for the adverse effects for mothers. The burden on continuously-operating hospitals is likely much less significant in our setting where the hospitals experiencing closures tend to be small relative to the absorbing hospitals. Additionally, the complier population differs in Sweden where mothers are assigned to a local delivery hospital, whereas in the United States mothers have more freedom over their choice. In a concurrent working paper, Battaglia (2022) examines maternity ward closures in the United States (1996 to 2018). Consistent with our own work, Battaglia (2022) estimates declines in C-sections and null effects on infant mortality. While Battaglia (2022) focuses mostly on the birth environment with respect to C-sections, we also analyze quality of care and quality-related outcomes such as maternal morbidity.

As the closures cause the diversion of women to nearby counties, this work also adds new insights in the role of geography in health care utilization (Wennberg and Gittelsohn, 1973; Baicker et al., 2006; Chandra and Staiger, 2007; Skinner, 2011; Finkelstein et al., 2016; Molitor, 2018; Deryugina and Molitor, 2020, 2021). Specifically relevant to perinatal care, this paper also augments discussions about the appropriate use of C-sections and the function of providers in that debate (Baicker et al., 2006; Currie and MacLeod, 2017).

2 Background on Closures

Hospital-based OB unit access has been on continual decline over the last 31 years (Figure 1). Panel A shows that the share of rural counties with an operational OB unit declined from 64% in 1989 to 43% in 2019. At the same time, rates of infant mortality in rural counties have deteriorated relative to urban counties. The closures are geographically diverse (Figure 1B), albeit more intense in states with more significant rural populations. Most states have at least one county with a closure during this time period.

Why are rural OB units closing? Closures are most commonly attributed to financial pressures resulting from uncompensated care and insufficient public payer reimbursements (Lindrooth et al., 2018; Kaufman et al., 2016; Zhao, 2007). Rural hospitals have disproportionately shouldered the burden of recent reductions in Medicaid and Medicare reimbursement rates as rural hospitals exhibit higher rates of Medicaid usage, elevated rates of poverty, and serve an aging population (Hung et al., 2016; Kozhimannil, 2014). For efficiency reasons, large hospital networks often consolidate operations by closing their financially struggling facilities —which tend to be smaller and more rural —and reallocate resources to their larger, more urban hospitals. Another contributing factor is staffing shortages driven by a declining supply of family physicians with OB training (Tong et al., 2012, 2013; Cohen and Coco, 2009; Zhao, 2007). It is also possible that demand-side factors such as demographic changes including a shrinking rural population along with an aging population have added to the pressure to close (Wishner et al., 2016).

OB units are dedicated hospital services that provide care to mothers and infants in the period leading up to birth (prenatal care) and at the time of birth (intrapartum care). OB unit closures may impact maternal and infant health through at least four channels. First, closures reduce proximity to prenatal care. Prenatal care includes routine ultrasound and blood tests, management of existing conditions, information for having a healthy pregnancy, and developing a birth plan. As there is (debated) evidence that prenatal care improves birth outcomes, closures may result in lower gestation lengths and, consequently, lower birth weights (Alexander and Korenbrot, 1995).

Second, closures reduce proximity to intrapartum care. Expecting mothers now must travel farther to give birth in a hospital. Increased travel distance at the time of labor could lead to worse outcomes if the travel time causes delays in receiving medical attention, or if it causes women to give birth in non-hospital settings. Third, closures could lead to crowding, negatively impacting outcomes if the remaining OB units become oversubscribed.

Each of the first three channels predict closures lead to worse outcomes. However, a fourth possibility is that closures may reallocate patients to a different type of hospital, thereby potentially changing the quality of care they receive at the time of birth. If OB units are closing in lower quality hospitals and those patients are redirected to higher quality hospitals, closures may improve outcomes.

3 Data

3.1 Birth-Related Outcomes

Our core data sources are the natality and mortality files from the National Vital Statistics System (NVSS) for 1989-2019. The natality (mortality) files cover the near universe of births (deaths) in the United States. Each observation in these data is a birth (death) and these data come from birth (death) certificates. The NVSS natality files include information on both the infant and parents, and include location of birth (e.g., county of occurrence; occurrence in a hospital), procedures (e.g., induction; C-section), and numerous measures of infant and maternal health. We construct composite measures of infant and maternal health to summarize the many outcomes. We use a restricted-access version of these files which include county of birth and county of residence identifiers. Many of natality variables are not available for the entire sample period as they were phased in or out with the rollout of the revised birth certificate beginning in 2003. To account for this, we construct one composite measure for infant health (available 1989-2006) and two composite measures for maternal health; one available 1989-2006, and another available 2009-2019. Details for the construction of these composite measures are provided in Section A.3. The NVSS mortality files allow us to examine infant mortality rates.

3.2 Identifying Closures

A "closure" is defined as the loss of all hospital-based OB units in a given county. We identify closures using two independent data sources and methods. In our preferred method, we use the NVSS natality files and infer a closure when the number of hospital-based births occurring in a county in a given year drops to near zero. See Sections A.1.1 and A.1.2 for more details on our algorithm for identifying closures. Using an alternative method, we rely on data from the American Hospital Association (AHA) Annual Surveys from 1995-2016 which reports operational hospital services by year. While the AHA data has the advantage of being hospital- rather than county-level, the survey nature of the data may induce measurement error. Nevertheless, both measures are largely in agreement. We report estimates for the main outcomes using the AHA-based coding in Table A1, which are similar to our preferred estimates albeit slightly less precise. Unless otherwise noted, we use the NVSS-based method of identifying closures throughout the paper.

We identify 605 counties that experienced the loss of all OB services at some point during our 31-year sample and the trend has been steady over this period. While OB services resumed in some of these counties, 488 counties experienced a closure without a subsequent reopening. There were 33 counties that experienced an opening without a prior closure.

3.3 Quality Metrics: Mechanisms

To understand mechanisms, we augment the NVSS natality files with data from the AHA Annual Surveys and Hospital Compare. Specifically, we merge each birth with county-level characteristics based on the county of birth. Using AHA Annual Surveys we proxy for OB resources with the presence of a neonatal intensive care unit (NICU). Using CMS Hospital Compare files, we measure general hospital quality using a composite of four standard quality metrics (process measures, patient satisfaction surveys, risk-adjusted readmission rates, and risk-adjusted mortality rates). More detail on the construction of these measures is provided in Section A.2, and summary statistics for all main outcomes can be found in Table A2.

4 Empirical Framework

We estimate the impacts of OB unit closures using a difference-in-differences (DD) design, which we implement using a two-way fixed effects (TWFE) specification:

$$\text{Y}_{cy} = \beta \text{Closed}_{cy} + \gamma \text{X}_{cy} + \delta_c + \delta_{uy} + \varepsilon_{cy} \tag{1}$$

In Eq. (1), Y_{cy} represents the outcome for mothers (infants) residing in county c, who give birth (are born) in year y. Our treatment variable, $Closed_{cy}$, is an indicator equal to one in the years following the loss of all hospital-based OB units in the mother's county of residence.[5] We analyze

[5] To ensure a "staggered" DD framework in which treatment turns on but not off, we drop 117 counties that experience both closures and openings (primarily reopening after closure) and 33 counties that experienced only an opening. We find similar results in alternative models that include these additional counties and allow the treatment status to change more than once (Table A3).

a comprehensive set of outcomes including the location of birth, several measures of infant and maternal health, and characteristics of hospitals in the county of birth occurrence. X_{cy} represent time-varying county level covariates: population shares for 5-year age bands, per-capita personal income, per-capita government transfers, and the employment-population ratio. δ_c are county fixed effects, which ensure the estimates are identified from variation within counties rather than cross-sectional comparisons. δ_{uy} are urban group-by-year fixed effects, which allow the idiosyncratic time effects to vary by the six groups in the (time-invariant) 2013 NCHS urban/rural coding.[6] These are potentially important given that the closures we analyze are mainly rural and time shocks may not be accurately captured by a single set of time fixed effects.

In order to interpret β as the causal effect of closures on health outcomes, the standard DD parallel trends assumption must hold. In this setting it requires that OB closures are uncorrelated with other unobserved time-varying determinants of maternal and infant health outcomes. An obvious concern is that closures are not randomly assigned across counties. For example, closure counties have smaller and less urban populations (Table A2). While county fixed effects account for cross-sectional time-invariant differences, it is possible that some of the forces determining closures (e.g., demographic shifts) induce differential trends in the outcomes between treated and untreated counties. The urban group-by-year fixed effects alleviate this concern to an extent, but we probe this concern further in three ways.

First, we conduct a series of balance tests in which we replace the outcome from Eq. (1) with the fertility rate and 15 maternal characteristics. The results for this test are presented in Figure A1 and reveal slight imbalance in three of the 16 variables (the three race variables). Second, to mitigate concerns about possible imbalance, we estimate each county's propensity to experience a closure using their 1989 characteristics then weight control observations based on this propensity. This gives more weight to rural counties and essentially zero weight to dense and highly populated urban counties (Section A.4.2). We find no evidence of imbalance when using these weights. Our main results are similar across weighted and unweighted specifications (Tables A4 to A6), suggesting any imbalance, if it exists, has minimal effects on our estimates. We also consider more parsimonious versions (e.g., excluding controls, using year fixed effects in place of δ_{uy}) and richer versions (e.g., including state-by-year fixed effects) of Eq. (1) for robustness.

Third, we present our main results in an event study framework; the details of the specification are discussed in Section A.4.3. While the balance tests suggest our specification sufficiently accounts for long-term demographic shifts on a set of observables, unobservable shifts could still be problematic. The event studies allow us to abstract from long-term trends (e.g., the factors discussed in Section 2) and observe whether changes in the outcomes coincide precisely with the

[6]Similar controls are used in Bailey and Goodman-Bacon (2015), who analyze the establishment of community health centers in mostly urban counties.

timing of treatment. The nature of the treatment is such that we expect the impacts to materialize immediately if the estimated relationship is causal.

TWFE approaches to DD designs can produce biased estimates when treatment effects are heterogenous (de Chaisemartin and D'Haultfoeuille, 2020; Goodman-Bacon, 2021). We present results from two alternative DD estimators addressing the negative weighting concern —the de Chaisemartin and D'Haultfoeuille (2020) estimator is presented alongside the main TWFE results and Borusyak et al. (2021) imputation-based event study estimates are provided in Figure A2. We discuss the issue in more detail in Section A.4.1 and show that this type of bias is minimal in our setting.

5 Results

5.1 Main Results

Figure 2 and Figure 3 present event studies and the corresponding average effects for the main outcomes. For most outcomes, there is little evidence of meaningful differences in outcomes between the treated and untreated counties in periods leading up to the closure. For a number of outcomes—those that are impacted significantly by closures—there is a statistically significant and discrete change in the outcome that coincides precisely with the timing of treatment. Overall, the event studies provide evidence in support of the parallel trends assumption, and lend credence to our causal interpretation.

We seek to understand how closures impact health through both prenatal and intrapartum care. Figure 2 explores how closures affect the location of birth, prenatal care, and outcomes that are mostly determined by the environment leading up to birth. Figure 3 addresses the impact of closures on more downstream outcomes, which are also affected by conditions at the time of birth (e.g., having to travel a long distance while in labor or the quality of the birth hospital).

Figure 2A reveals that when a mother loses the remaining OB unit in her county of residence, the probability of giving birth in her county of residence declines by 30 percentage points (pp) on a base of approximately 30%. Figure 2B reveals a statistically significant decline in the share of births occurring in a hospital, though the magnitude is miniscule (-0.2pp on a base of 98.7%). Together, Figure 2A and B indicate that after a closure occurs, nearly all births are diverted to hospitals in other counties rather than leading to a large number of out-of-hospital births.

Figure 2C confirms that closures reduce access to prenatal care and reveals a small but statistically significant decrease in the number of prenatal visits (1.5% decline). Prenatal care has long been associated with healthier birthweight and gestational age, though the causal link is less clear (Alexander and Korenbrot, 1995). Given our documented effect of closures on prenatal care, it is natural to ask whether these birth outcomes deteriorate. Figure 2D and E show that closures lead to a statistically insignificant decline in birthweight (-2.18 grams, p-value=0.221), but a significant decline in gestational age (-0.045 weeks, p-value<0.001). It is possible this effect on gestational age is driven by an increase in premature births, a severe outcome, or alternatively by slightly early births which would be less concerning. Figure 2F shows no impact of closures on premature births,[7] while Figure 2G shows that the gestation effect is driven by an increase in births between 37 and 39 weeks (1pp, p-value<0.001).

On the surface one might conclude the decline in prenatal visits causes the documented rise in births at 37-39 weeks. However, Figure 2H reveals that inductions at 37-39 weeks increase by approximately the same magnitude (1pp, p-value<0.001). Hence, the entire effect on early births can be explained by increased inductions. For both births and inductions at 37-39 weeks, about two-thirds of the increase is due to increased births and inductions specifically at 39 weeks (Table A7). It is likely that providers schedule inductions to avoid long travel at the time of naturally occurring labor.[8] In conclusion, Figure 2 suggests that the rise in early births is a consequence of more scheduled births, as evidenced by the increase in inductions, and not necessarily because of complications that arise due to missed prenatal visits.

Figure 3 examines six outcomes that are a function of conditions at the time of labor and delivery. We have already shown that nearly all affected mothers travel to a hospital in another county to give birth. Travel itself may have direct negative consequences for maternal and infant health if it prevents a mother from obtaining medical attention within the appropriate timeframe of labor and delivery. On the other hand, closures also divert mothers to different hospitals, which could be welfare-improving if the receiving hospital is of higher quality. We examine C-sections —a common obstetric procedure, but an outcome that lacks obvious welfare implications —and five measures of infant and maternal health which have more straightforward welfare implications.

Figure 3A shows that closures lead to a clear and substantial decline in C-sections (-1.1pp, p-value<0.001). We unpack the mechanisms underlying this effect and attempt to draw welfare implications in the following section.

Figure 3B presents estimates for the most severe outcome: infant mortality. The TWFE estimate yields a null effect on infant mortality, and the confidence interval allows us to rule out a relatively small increase (7%) in infant mortality. While the TWFE estimate yields a null effect and the event study reveals no change in the outcome coinciding with the timing of treatment, the de Chaisemartin and D'Haultfoeuille (2020) estimate is positive and significant. However, this appears to be anomalous. The de Chaisemartin and D'Haultfoeuille (2020) average effect esti-

[7] Similarly, there are no impacts on the share of births with low or very low birthweight (Table A4).

[8] The American College of Obstetricians and Gynecologists lists living far from the hospital as a reason to consider elective induction at 39 weeks. See https://www.acog.org/womens-health/faqs/labor-induction.

mator is calculated using only the period prior to treatment ($t = -1$) as the comparison period whereas the TWFE estimator uses the entire pre-treatment period. An idiosyncratic drop in the outcome at period $t = -1$ therefore yields a positive effect. A version of the de Chaisemartin and D'Haultfoeuille (2020) estimator using period $t = -2$ as the comparison group would yield a null effect. Various alternative specifications also yield no effect (Table A5).[9]

Figure 3C and D present results for two less severe measures of infant health. Figure 3C analyzes the share of infants with low Apgar scores, a standard infant health measure available for the entire sample period. Figure 3D analyzes a composite measure of infant morbidity composed of variables available in state-years using the unrevised birth certificates (see figure notes for more details on composite measures). Neither outcome reveals a significant impact of closures on infant health. Figure 3E and F present estimates for two composite measures of maternal morbidity. Figure 3E uses a set of variables available in state-years using unrevised birth certificates while Figure 3F uses a more comprehensive set of maternal morbidity measures that were introduced in 2009 for states using revised birth certificates. While we observe no impact on the unrevised measure, there is a robust 0.1 standard deviation decrease (improvement) in the revised maternal morbidity measure coinciding precisely with the timing of treatment (p-value = 0.001). The improvement in maternal morbidity can largely be attributed to reductions in maternal blood transfusions and perineal lacerations (Figure A3 provides estimates for all components of the composite measures). Overall, Figure 3 suggests that the average effects of closures on welfare-relevant measures of health are either negligible or slightly beneficial.

5.2 Mechanisms

We next explore possible mechanisms underlying the average impacts of closures. To begin, we focus on understanding the significant (1.1pp) decrease in C-sections. There are at least two possible channels underlying this decline. First, a recent randomized-controlled trial found that induction at 39 weeks (as opposed to expectant management) decreases the probability of C-section by 16% (Grobman et al., 2018). As such, if women in counties experiencing closures are more likely to have a scheduled induction to avoid travel during labor, then it is likely that C-sections would decrease. Overall, we find that closures increase the probability of induction by 1.8pp (Table A5). Using the estimate from Grobman et al. (2018), this implies a reduction in C-sections of 0.3pp, explaining about one third of the overall effect.

A second possible mechanism is that women are reallocated to hospitals with different C-section practices. To explore this possibility, Figure 4A tests whether closures induce women to

[9] In addition to varying the covariates and fixed effects, Table A5 also presents results for neonatal mortality (death within 28 days of birth). Compared to infant mortality, the estimate for neonatal mortality is closer to zero (coefficient = 0.0173) and more precisely estimated (standard error = 0.141).

give birth in counties with different C-section rates. To ensure the outcome is not mechanically related to changes in a mother's own propensity to have a C-section, the outcome for each mother residing in a closure county is the risk-adjusted C-section rate in her county of birth occurrence in the three years prior to closure (for mothers residing in non-closure counties, it is a random three year period).[10] As such, changes in the outcome derive only from mothers changing where they give birth, rather than changes in their own propensity to have a C-section.

Figure 4A shows that closures prompt women to give birth in counties that have, on average, 1pp lower risk-adjusted C-section rates. In Figure 4B we investigate the extent to which this reduction in local C-section rates influences a mother's own probability of C-section. We find that while on average mothers are reallocated to counties with lower C-section rates, there is substantial heterogeneity across the large number of closures. We document this heterogeneity by calculating for each closure, the pre-closure gap in risk-adjusted C-section rates between the closure county and the counties in which mothers are most likely to give birth post-closure (the "receiving" county). Specifically, the receiving county is defined as a weighted average of all counties with any pre-closure market share among mothers residing in the closure county, weighted by their market share. We caution against causal interpretations of these heterogeneous effects due to lack of exogenous variation in C-section gaps.

Figure 4B plots the distribution of C-section gaps across all closures. While the center of the distribution is negative (median = -0.027; mean = -0.034) as expected given the results from Figure 4A, there is mass on both sides of zero and substantial variation overall. If local C-section rates are an important determinant of a mother's own probability of C-section, then the effect of closures on C-sections should be heterogeneous with respect to the C-section gaps. To test this, we estimate whether the impact of a closure on C-sections is different for closures above and below the median C-section gap.[11] Figure 4B reports these estimates and reveals that the effect of a closure on C-sections is particularly large (-2.1pp, p-value$<$0.001) for closures that induce mothers to give birth in counties with much lower C-section rates (i.e., below median C-section gap). The differential effect of being above the median relative to below is significant (1.4pp, p-value=0.001), confirming that local C-section rates are likely an important determinant of a mother's probability of C-section. In summary, reallocation to hospitals with lower C-section rates is likely the dominant mechanism explaining the overall decline in C-sections.

C-sections are widely considered to be overused; as such, it is tempting to view the estimated

[10] C-section rates are risk-adjusted to account for differences in patient mix between closure and non-closure counties.

[11] This is operationalized via estimating a version of Eq. (1) that also includes the closure indicator interacted with an indicator for above the median C-section gap, where the outcome is the share of births delivered via C-section (i.e., the same outcome as Figure 3A).

decrease in C-sections as welfare-improving.[12] However, Currie and MacLeod (2017) show that health outcomes improve when C-section rates are either: decreased among mothers with a low predicted risk of C-section, or increased among mothers at high-risk. Thus, it is reasonable to conclude that the decreases we observe in C-sections are welfare-improving only if they are concentrated among low-risk women. Following Currie and MacLeod (2017), we predict the probability of C-section using the full sample of individual-level data and a range of risk factors, and estimate the effects of closures on C-sections across quartiles of the risk distribution. We find statistically significant declines in C-sections among all four quartiles (Figure A4).[13] These findings highlight the difficulty in drawing welfare conclusions from reductions in C-sections alone.

Next we focus on uncovering the mechanism for the remaining outcomes presented in Figure 3, the morbidity and mortality measures. Unlike C-sections, changes in these outcomes have clear welfare implications. Recall that there are four likely mechanisms through which closures could affect health: (1) increased travel during labor, (2) OB unit crowding in the remaining units, (3) reduced prenatal care, and (4) reallocation to higher quality hospitals. Travel, crowding, and reduced prenatal care are channels that would explain negative health impacts of closures, while a reallocation channel would likely produce better health outcomes. It is possible that any of these channels are at work (with the harmful and beneficial channels competing), but since we find closures have null or slightly beneficial effects on infant and maternal health, this suggests reallocation to higher quality hospitals is the dominant mechanism.

While our focus is on the reallocation mechanism, we investigate other possibilities as well. Figure A5 provides estimates from a version of Eq. (1) that replaces the closure indicator with a quadratic in distance to the nearest OB unit in order to test whether deleterious impacts appear at longer distances. The nonlinear estimates are in strong agreement with the main estimates, and we find no negative impacts emerging at longer distances. We also note that the crowding mechanism is unlikely an important factor in our setting: in the pre-closure period, the number of births in closure counties was only 3% of the number of births in receiving counties.

In Figure 5, we test whether the average closure diverts women to higher quality hospitals. In each plot, the outcome is defined as a measure of hospital quality in each mother's county of birth occurrence. We measure hospital quality in two ways. First, we use a general measure of quality from Hospital Compare. Hospital Compare provides several quality measures, and ours is a composite of four commonly used measures (Doyle et al., 2019).[14] Second, we measure OB-

[12] Reducing C-sections among low-risk women is a target of the Healthy People 2030 objectives.

[13] See the figure notes for more details on risk prediction.

[14] Details on the construction of these measures can be found in Section A.2.1, and estimates for each of the components of the composite can be found in Figure A3.

specific hospital resources using the presence of a NICU.[15] Both Figure 5A and Figure 5B provide clear evidence that closures, on average, prompt women to give birth in counties with higher quality hospitals. Figure 5A shows closures lead women to give birth in counties that have 0.1 standard deviations higher quality scores, and Figure 5B reveals that women are 4pp more likely to give birth in a county with a NICU.

In Figure 5C and Figure 5D, we replicate the exercise of plotting pre-closure gaps (shown in Figure 4B), but for our two measures of hospital quality. In comparison to the C-section gaps, the quality gaps are overwhelmingly positive. That is, nearly all closures reallocate women to counties with higher quality hospitals.

We next focus on the health outcome for which we find a statistically significant improvement on average, the revised maternal morbidity composite (Figure 3F), and test whether the effects of closures are heterogeneous across the distribution of quality gaps. Figure 5C shows that the closure effect is essentially identical above and below the median Hospital Compare quality gap. This may reflect the fact that this measure is only a noisy proxy of true hospital quality (i.e., much of the heterogeneity could be due to noise). It is also possible that this metric does not adequately measure the relevant dimension of quality, as it is not specific to obstetrics. Alternatively, it could be indicative of no effect of quality on maternal morbidity. Figure 5D relies on an OB-specific proxy for quality, the presence of a NICU. We find that the effect of closures on improvements in maternal morbidity is particularly large (0.122 standard deviation decrease) and statistically significant for closures that are most likely to induce mothers to give birth in a county with a NICU (i.e., NICU gap is above median). While this estimate is nearly twice as large as the effect for closures with a below-median NICU gap (0.067 standard deviation decrease), the difference is not statistically significant. In conclusion, it is difficult to precisely measure the returns to quality given that health care quality measures tend to be noisy proxies. Despite this, we find compelling evidence that closures prompt mothers to give birth in counties with higher quality hospitals and more OB resources, suggesting that reallocation to better hospitals is a mechanism underlying the observed improvement in maternal health.

Policy and media discussions around closures often focus on the most severe outcomes. As such, lastly, we aim to unpack the null effect on infant mortality. For any of the mechanisms we have discussed to drive changes in infant mortality, it must be the case that closures change the behavior of mothers whose infants are at high risk of death. If high-risk mothers are not treatment compliers, this could explain the null effect. To assess this possibility, we utilize the linked birth-infant death data to construct a predicted probability of infant death for every birth. We then estimate a "first-stage" regression (outcome is birth in county of residence, as in Figure 2A)

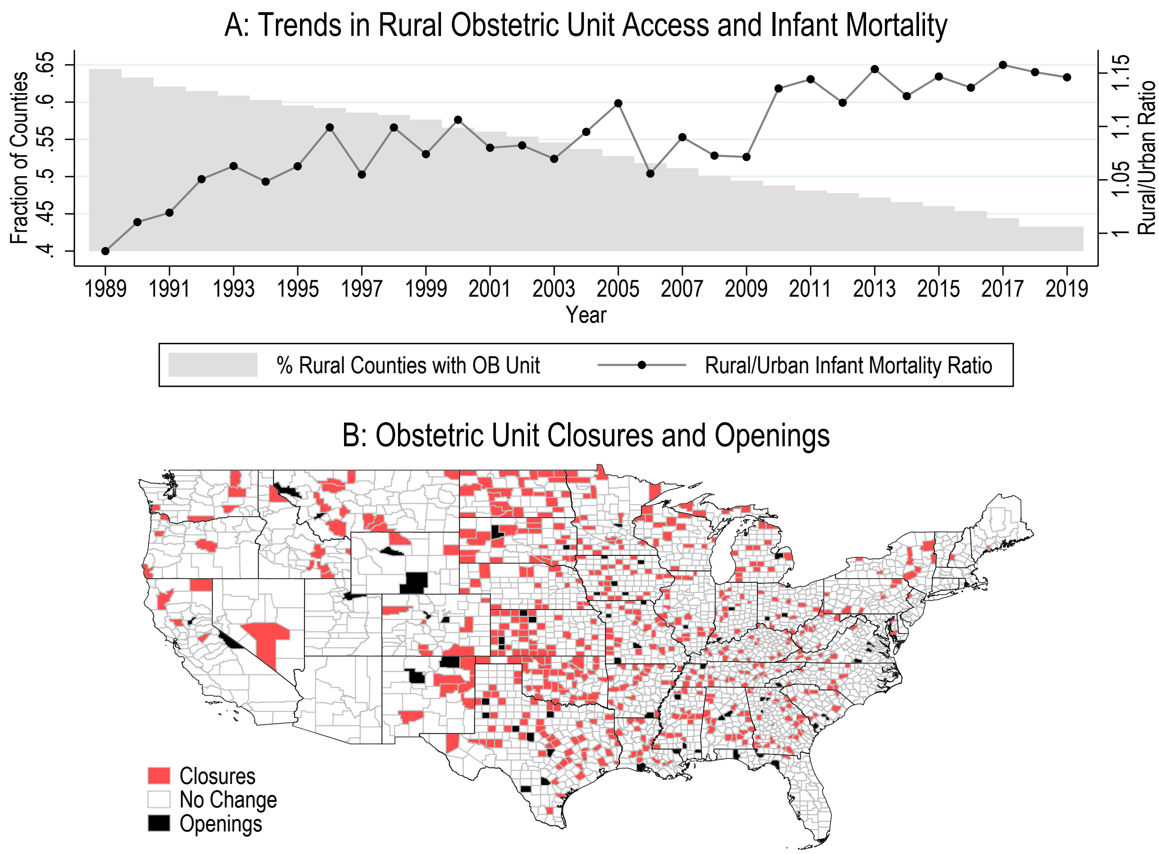
[15] We determine whether there is an operational NICU in each county using data from the AHA. We focus specifically on large NICUs (>25 beds), as Phibbs et al. (2007) show high-volume NICUs are more effective.

across risk groups. Figure A6 plots these estimates across vigintiles (plus >99th percentile) of infant mortality risk and shows that mothers with the observably highest risk pregnancies (>99th percentile) are half as likely to be compliers compared to the average birth. This implies that high-risk pregnant mothers were already traveling outside of their county prior to the closures to give birth. Consequently, as the complier mothers are less likely to have complicated deliveries, we should not expect closures to affect extreme outcomes, such as infant mortality.

6 Conclusion

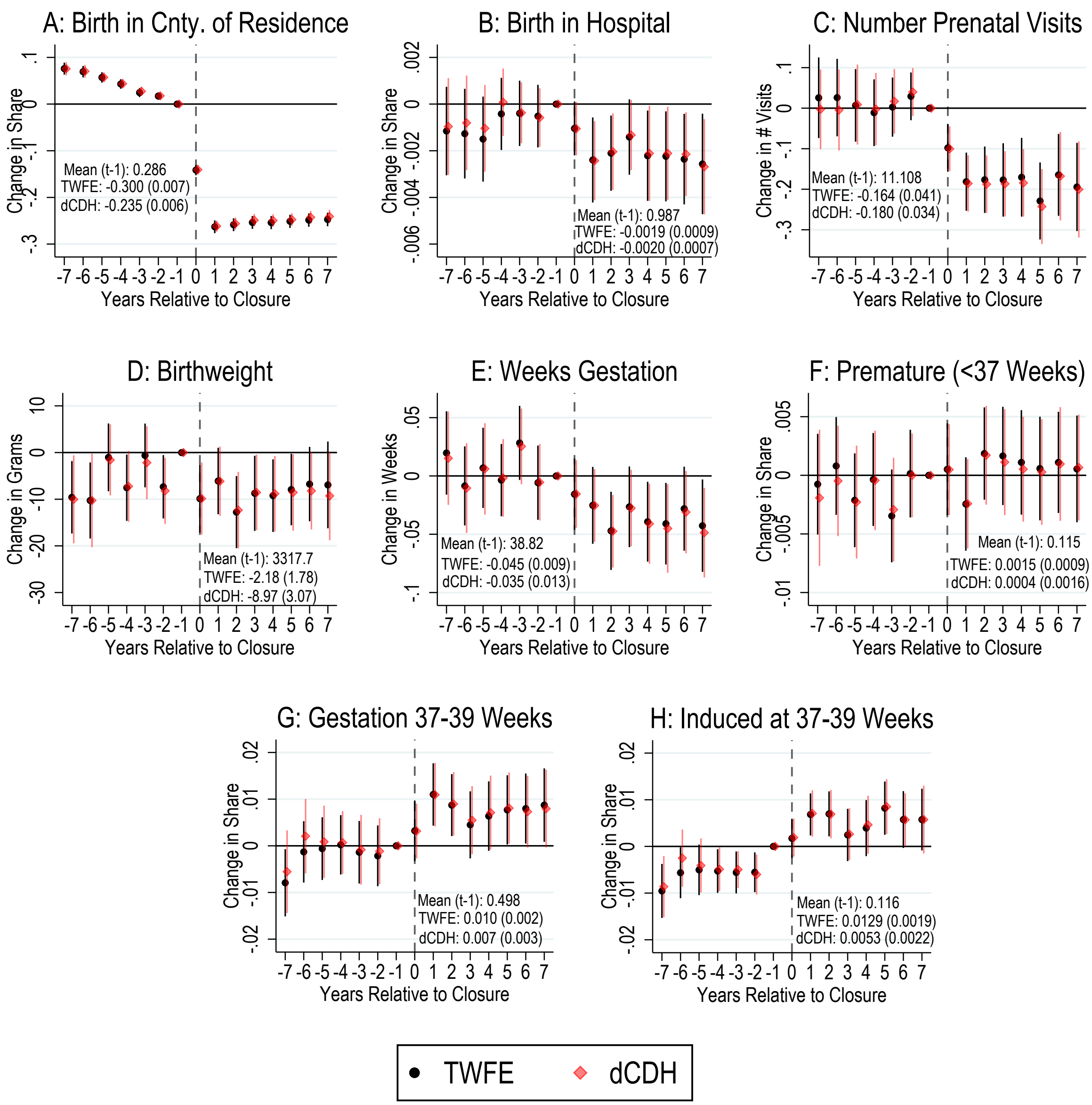
The trend in the regionalization of perinatal health care has left many counties in the United States without a hospital-based OB unit. At the same time, rates of infant mortality in rural counties relative to urban counties have been steadily increasing —causing concern that these two phenomena may be linked. Studying the closures of obstetric units across three decades, we conclude that the closures, as best we can measure, do not lead to worse health outcomes for mothers and their infants. While many mothers must travel farther for care, they receive care at better equipped hospitals. These receiving hospitals also perform fewer C-sections, and consequently, lead impacted mothers to have fewer C-sections themselves —emphasizing the strong role of place-based effects in health care.

Figure 1: Trends in Obstetric Unit Access and Infant Mortality: 1989-2019



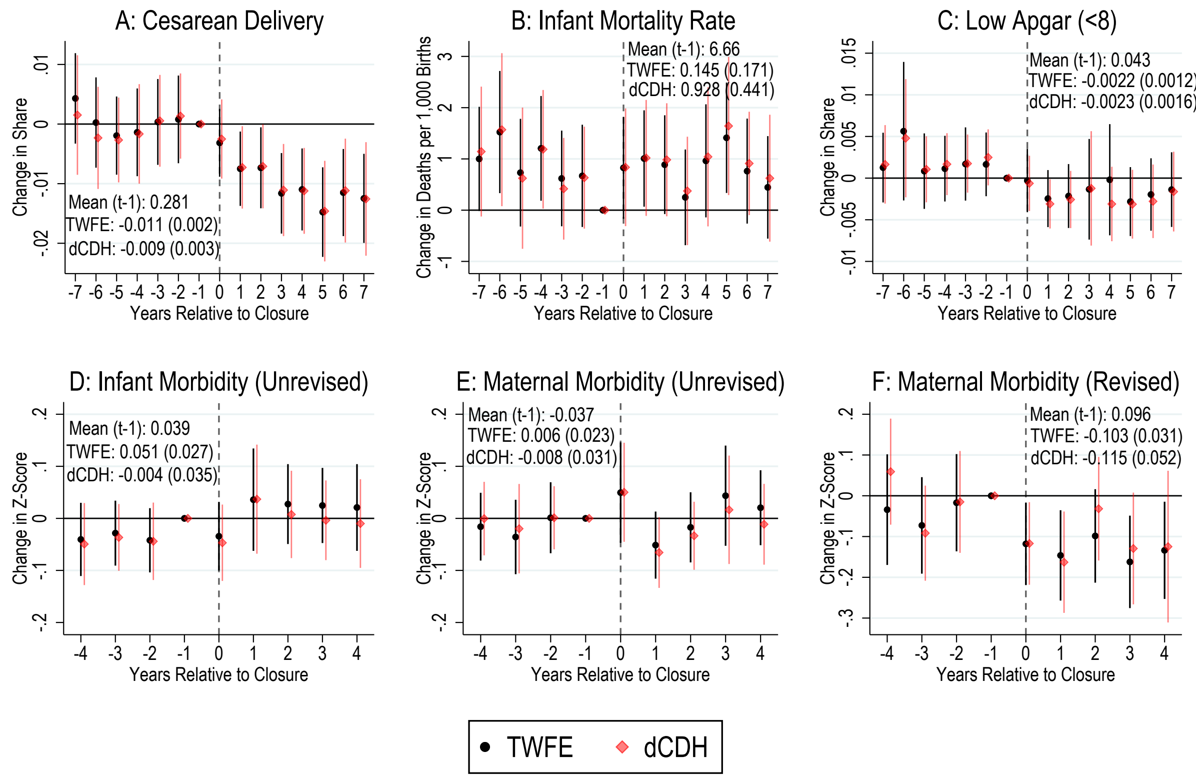
Notes: Rural counties are those classified as non-core or micropolitan in the 2013 NCHS urban/rural classification. In our sample, there are 1,883 rural counties and 1,065 urban counties. In Panel A, the shaded region displays the share of rural counties with an operational maternity ward in each year. The black line represents the infant mortality rate (IMR) in rural counties divided by the IMR in urban counties. In Panel B, a "closure" is defined as going from at least one operational maternity ward to zero, and an "opening" is the opposite. Counties shaded red (black) are those that experienced a closure (opening) in any year, and remained closed (open) through 2019.

Figure 2: Average Effect of Closures on Birth Location, Prenatal Care, and Outcomes Determined Prior to Delivery Experience



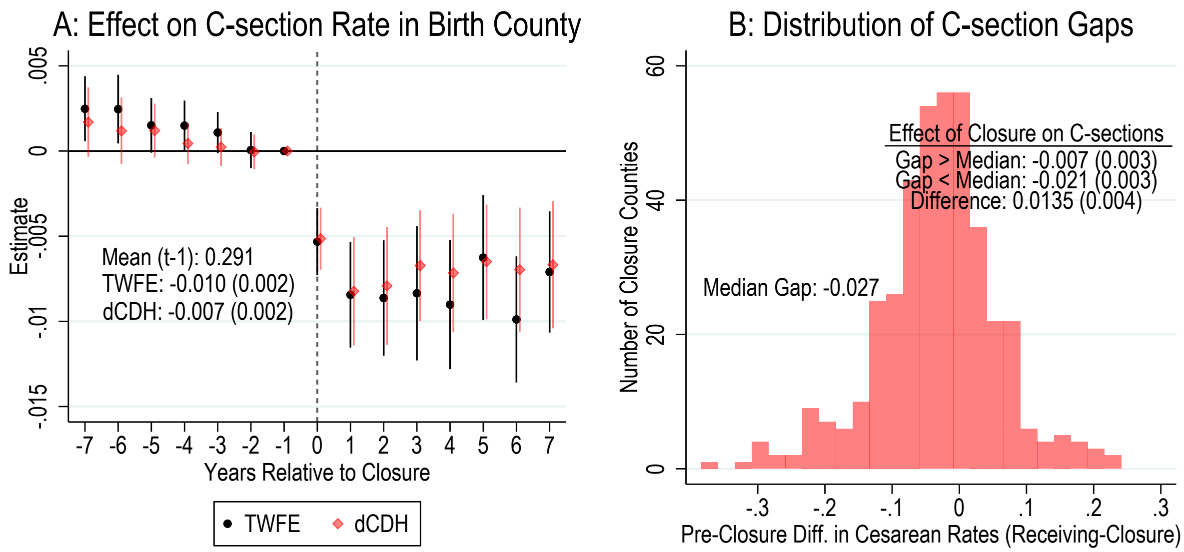
Notes: These figures plot estimated β_j from Eq. (2) using two different estimators. "TWFE" refers to estimates from a two-way fixed effects specification and "dCDH" refers to estimates from the de Chaisemartin and D'Haultfoeuille (2020) difference-in-differences estimator. Dynamic treatment effects are shown in black circles (TWFE) and red diamonds (dCDH), and bars represent 95% confidence intervals with standard errors clustered on county for both estimators. Each subfigure also displays mean of the dependent variable for treated counties in the year prior to closure, and the average treatment effects (with standard errors clustered on county in parentheses) for both estimators (i.e., estimates of β in Eq. (1)).

Figure 3: Average Effect of Closures on Infant and Maternal Health



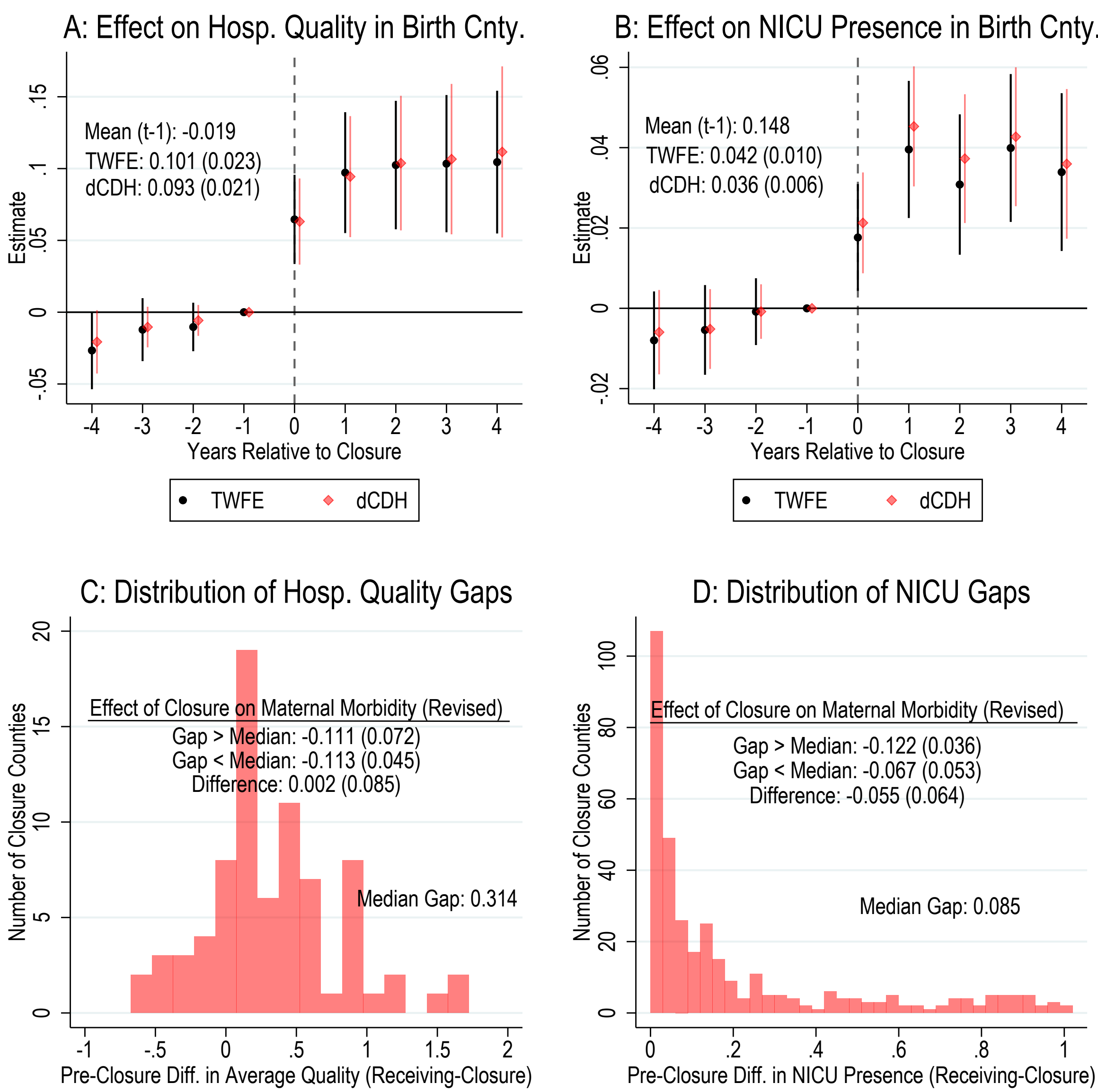
Notes: See Figure 2 for general notes on interpreting event studies. The three morbidity measures are composite outcomes that measure infant or maternal morbidity, where higher values represent worse health. Several components of these composite measures were phased in or out beginning with the 2003 revision of the birth certificate, and thus separate measures were created for state-years using unrevised or revised birth certificates. Each composite measure is limited to the states and years in which all components of the measure were available. "Infant Morbidity (Unrevised)" is available for 1989-2006 and is made of the following components: meconium staining, birth injury, infant seizures, and use of ventilator. "Maternal Morbidity (Unrevised)" is available for 1989-2006 and is made of the following components: maternal fever, excessive bleeding, and maternal seizures. "Maternal Morbidity (Revised)" is available for 2009-2019 and is made of the following components: maternal transfusion, 3rd-4th degree perineal laceration, ruptured uterus, unplanned hysterectomy, and admission to the ICU. More details on the construction of these measures and estimates for each component of the composite measures are provided in Section A.3. Because the composite measures use limited samples, the event studies are limited to four years pre- and post-closure.

Figure 4: Effect of Closures on Birth Environment (C-Section Delivery)



Notes: See Figure 2 for general notes on interpreting the event studies. In Panel A, the outcome for each mother is the risk-adjusted C-section rate in her county of birth occurrence in the three years prior to closure (for non-closure counties, it is a random three year period). Panel B displays the distribution of pre-closure C-section delivery gaps across all closure counties. For each closure county, we calculate the risk-adjusted C-section delivery rate in the three years prior to closure for births occurring in both the "receiving" and closure counties and the gap is the difference between these. The receiving county is defined as a weighted average of all counties with any pre-closure market share among mothers residing in the closure county, weighted by their market share. The text labelled "Effect of Closure on Cesarean Delivery" reports estimates from a version of Eq. (1) that includes an interaction term for the C-section gap being above median, and where the outcome is C-section delivery (i.e., the outcome from Figure 3A; not the outcome from Figure 4A). Sample restrictions for this analysis are discussed in Section A.4.4.

Figure 5: Effect of Closures on Birth Environment (Hospital Quality)



Notes: See Figure 2 for general notes on interpreting the event studies. "Hospital Quality" is a composite of four general hospital quality measures from Hospital Compare (processes of care, patient survey, risk-adjusted readmissions, and risk-adjusted mortality). Hospital Compare data are available beginning in 2010, thus all analyses of these data are limited to 2010-2019. "NICU" measures whether a NICU was operational in the county of birth occurrence. More details on the construction of these metrics, data sources, and estimates for the components of the Hospital Compare composite are provided in Section A.2. Panel B displays the distribution of pre-closure hospital quality (NICU) gaps between the closure and receiving counties. The text labelled "Effect of Closure on Maternal Morbidity (Revised)" reports estimates from a version of Eq. (1) that includes an interaction term for the hospital quality (NICU) gap being above median, and where the outcome is the revised maternal morbidity composite variable.

References

Alexander, Greg R and Carol C Korenbrot, "The role of prenatal care in preventing low birth weight," *The Future of Children*, 1995, pp. 103–120.

Avdic, Daniel, Petter Lundborg, and Johan Vikström, "How do closures of maternity wards affect birth outcomes," *Available at SSRN 3249886*, 2020.

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

Bailey, Martha J and Andrew Goodman-Bacon, "The War on Poverty's experiment in public medicine: Community health centers and the mortality of older Americans," *American Economic Review*, 2015, *105* (3), 1067–1104.

Battaglia, Emily, "The effect of hospital closures on maternal and infant health," 2022.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting event study designs: Robust and efficient estimation," *arXiv preprint arXiv:2108.12419*, 2021.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson, "Health care exceptionalism? Performance and allocation in the US health care sector," *American Economic Review*, 2016, *106* (8), 2110–44.

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

Cohen, Donna and Andrew Coco, "Declining trends in the provision of prenatal care visits by family physicians," *The Annals of Family Medicine*, 2009, *7* (2), 128–133.

Commonwealth, "The Rural Maternity Care Crisis," Technical Report, Commonwealth Fund 2019.

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

de Chaisemartin, Clément and Xavier D'Haultfoeuille, "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 2020, *110* (9), 2964–96.

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

_ **and** _ , "The causal effects of place on health and longevity," *Journal of Economic Perspectives*, 2021, *35* (4), 147–70.

Doyle, Joseph, John Graves, and Jonathan Gruber, "Evaluating measures of hospital quality: evidence from ambulance referral patterns," *Review of Economics and Statistics*, 2019, *101* (5), 841–852.

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

Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 2021.

Grobman, William A, Madeline M Rice, Uma M Reddy, Alan TN Tita, Robert M Silver, Gail Mallett, Kim Hill, Elizabeth A Thom, Yasser Y El-Sayed, Annette Perez-Delboy et al., "Labor induction versus expectant management in low-risk nulliparous women," *New England Journal of Medicine*, 2018, *379* (6), 513–523.

Hung, Peiyin, Carrie E Henning-Smith, Michelle M Casey, and Katy B Kozhimannil, "Access to obstetric services in rural counties still declining, with 9 percent losing services, 2004–14," *Health Affairs*, 2017, *36* (9), 1663–1671.

_ , Katy B Kozhimannil, Michelle M Casey, and Ira S Moscovice, "Why are obstetric units in rural hospitals closing their doors?," *Health Services Research*, 2016, *51* (4), 1546–1560.

Kaufman, Brystana G, Sharita R Thomas, Randy K Randolph, Julie R Perry, Kristie W Thompson, George M Holmes, and George H Pink, "The rising rate of rural hospital closures," *The Journal of Rural Health*, 2016, *32* (1), 35–43.

Kozhimannil, Katy, "Rural-urban differences in childbirth care, 2002-2010, and implications for the future," *Medical Care*, 2014, *52* (1), 4.

Kozhimannil, Katy B, Peiyin Hung, Carrie Henning-Smith, Michelle M Casey, and Shailendra Prasad, "Association between loss of hospital-based obstetric services and birth outcomes in rural counties in the United States," *JAMA*, 2018, *319* (12), 1239–1247.

Lindrooth, Richard C, Marcelo C Perraillon, Rose Y Hardy, and Gregory J Tung, "Understanding the relationship between Medicaid expansions and hospital closures," *Health Affairs*, 2018, *37* (1), 111–120.

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

Phibbs, Ciaran S, Laurence C Baker, Aaron B Caughey, Beate Danielsen, Susan K Schmitt, and Roderic H Phibbs, "Level and volume of neonatal intensive care and mortality in very-low-birth-weight infants," *New England Journal of Medicine*, 2007, *356* (21), 2165–2175.

Skinner, Jonathan, "Causes and consequences of regional variations in health care," in "Handbook of Health Economics," Vol. 2, Elsevier, 2011, pp. 45–93.

Tong, Sebastian T, Laura A Makaroff, Imam M Xierali, James C Puffer, Warren P Newton, and Andrew W Bazemore, "Family physicians in the maternity care workforce: factors influencing declining trends," *Maternal and Child Health Journal*, 2013, *17* (9), 1576–1581.

Tong, Sebastian TC, Laura A Makaroff, Imam M Xierali, Parwen Parhat, James C Puffer, Warren P Newton, and Andrew W Bazemore, "Proportion of family physicians providing maternity care continues to decline," *The Journal of the American Board of Family Medicine*, 2012, *25* (3), 270–271.

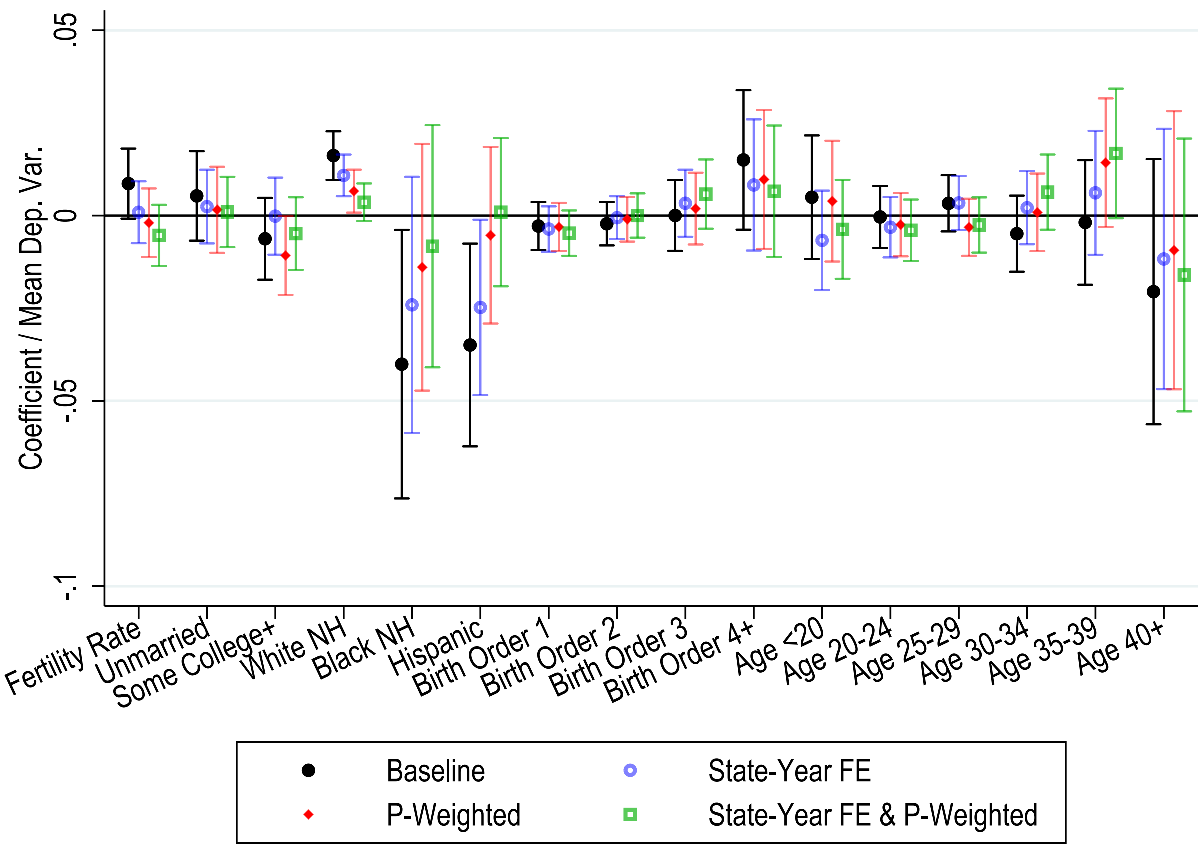
Wennberg, John and Alan Gittelsohn, "Small area variations in health care delivery: a population-based health information system can guide planning and regulatory decision-making.," *Science*, 1973, *182* (4117), 1102–1108.

Wishner, Jane, Patricia Solleveld, Robin Rudowitz, Julia Paradise, Larisa Antonisse et al., "A look at rural hospital closures and implications for access to care: three case studies," *Kaiser Family Foundation [Internet]*, 2016.

Zhao, Lan, *Why are fewer hospitals in the delivery business?*, NORC at the University of Chicago Bethesda (MD), 2007.

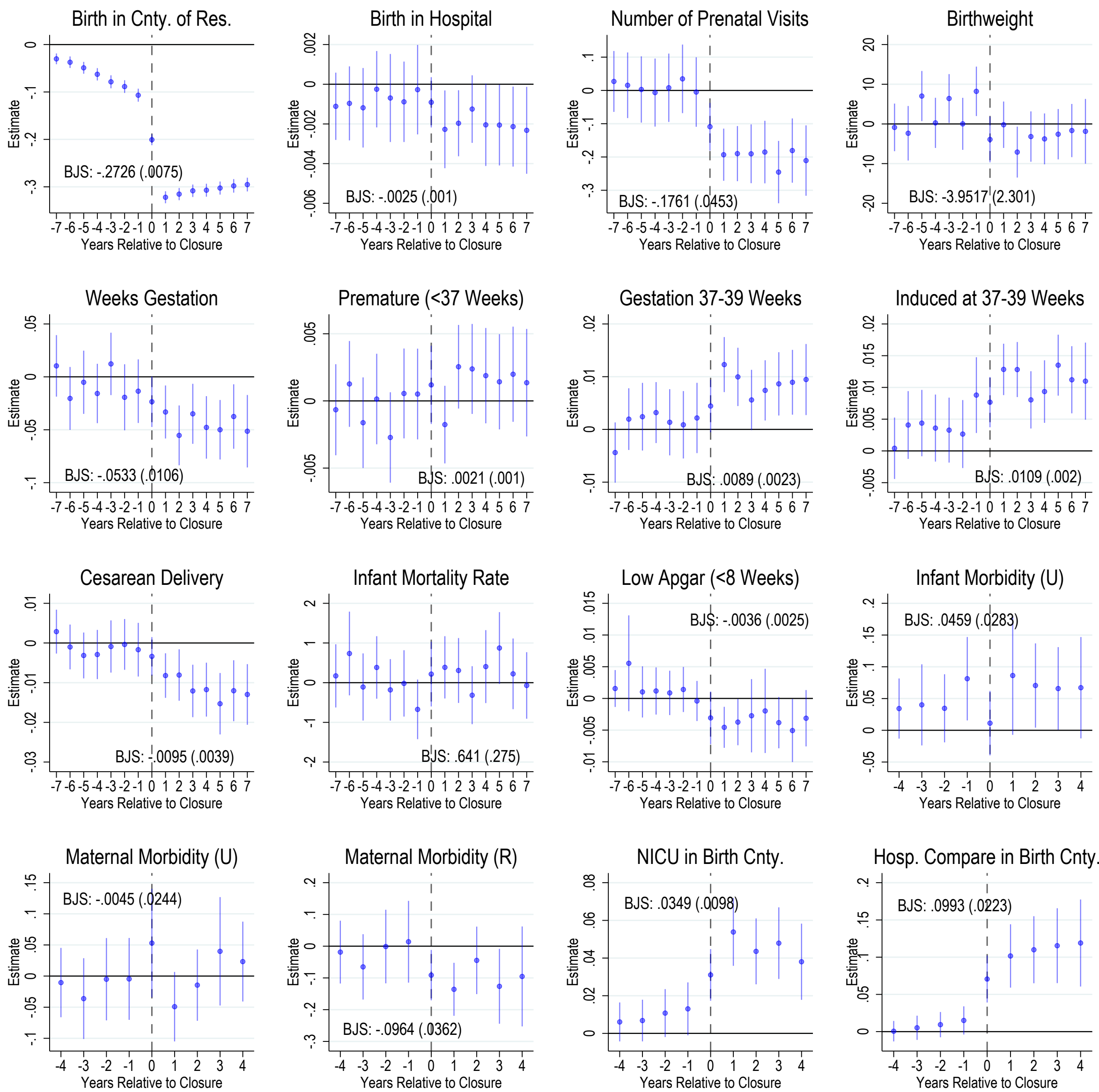
Appendix Figures

Figure A1: Effect of Closures on Fertility Rate and Mother Characteristics (Balance Test)



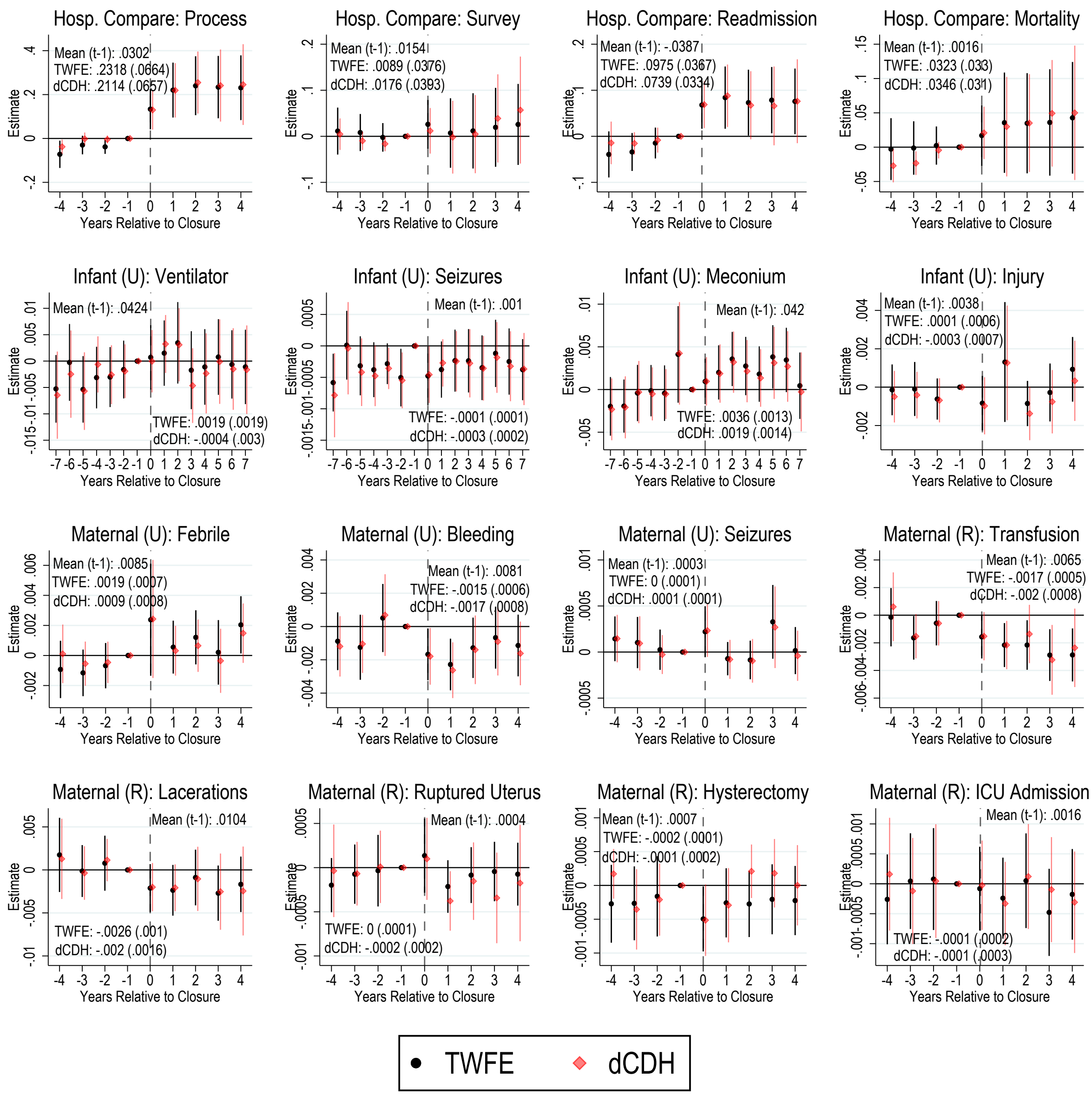
Notes: For comparability across outcomes, all coefficient estimates are divided by the mean of the dependent variable. The baseline specification (black) is described in Eq. (1). The second specification (blue) adds state-by-year fixed effects. The third specification (red) weights by the propensity to experience a closure. The process of calculating propensity score weights is described in Section A.4.2. Note that the weighted regressions are not balanced by construction: these regressions test for changes in these characteristics whereas the propensity weights are constructed from a cross-sectional logit. Furthermore, the weights are constructed based on a set of county-level characteristics rather than these mother characteristics. The fourth specification (green) includes state-by-year fixed effects and propensity weights.

Figure A2: Effect of Closures using Borusyak et al. (2021) Estimator



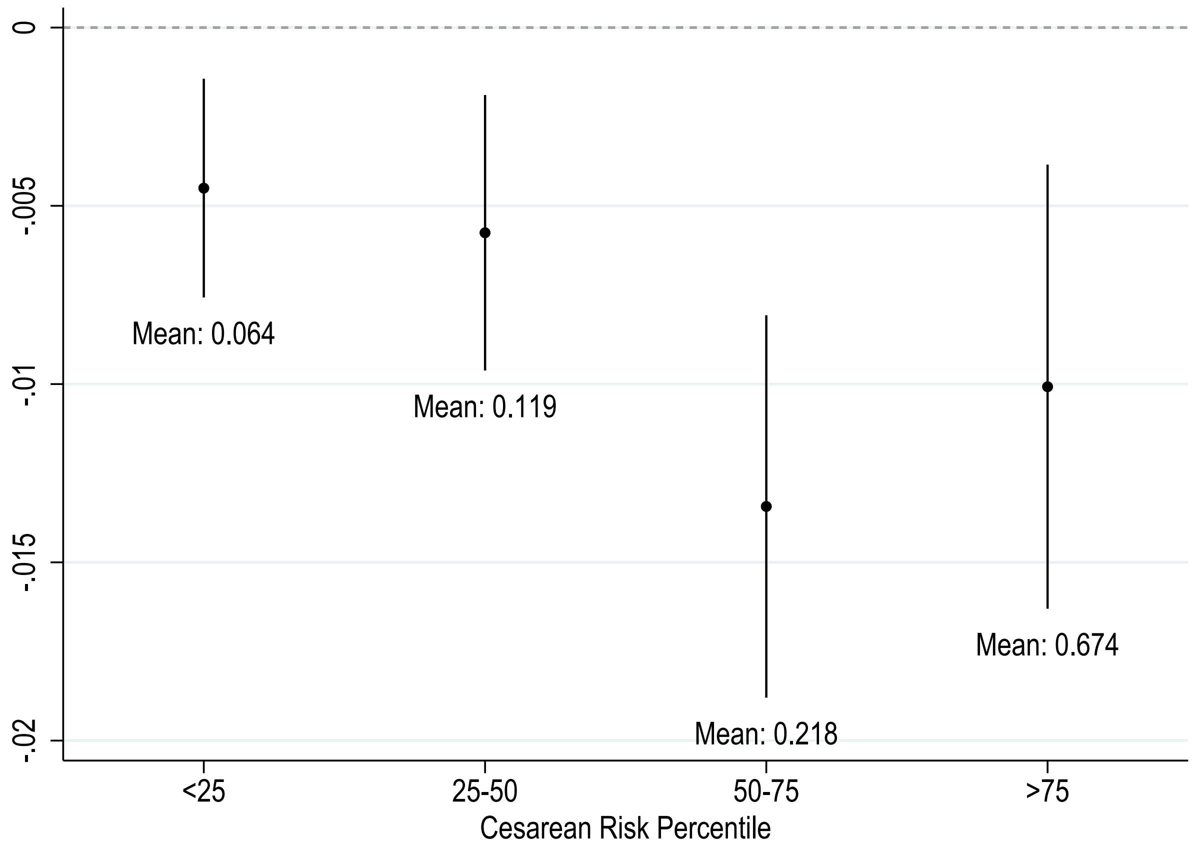
Notes: These plots replicate the estimates presented in Figure 2, Figure 3, and Figure 4 using the Borusyak et al. (2021) imputation-based difference-in-differences estimator. The point estimate labelled "BJS" on each plot represents the average effect in the entire post-treatment period. All estimates use the main specification, which includes controls for age-specific population shares and economic controls (employment-population ratio, per capita income, per capita transfers) and urban group-by-year fixed effects. The infant and maternal morbidity outcomes are composite measures, where "U" represents measures from the unrevised birth certificates and "R" represents measures from the revised birth certificates.

Figure A3: Effect of Closures on Components of Composite Measures



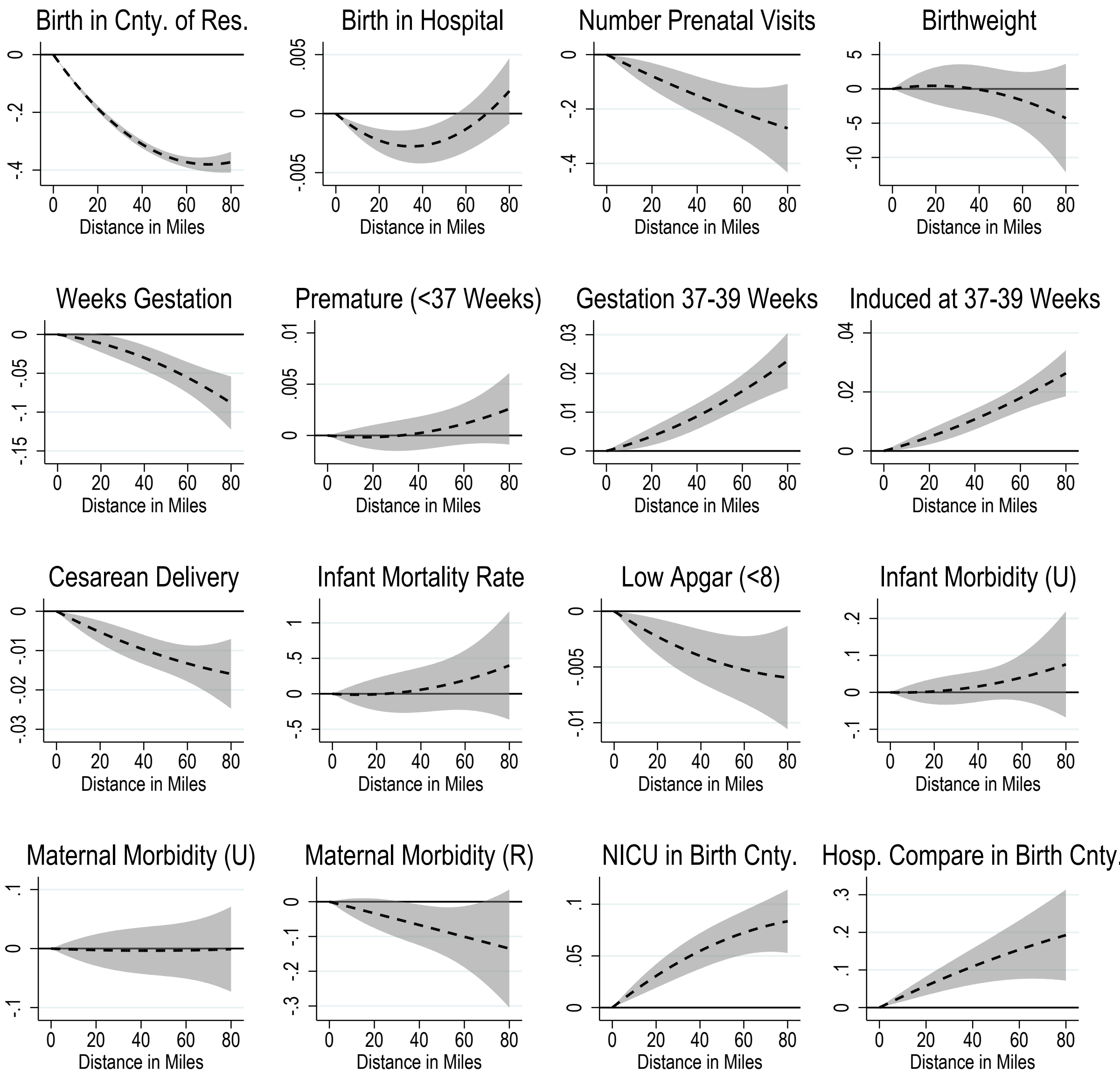
Notes: See Figure 2 for general notes on interpreting event studies. The top four plots show effects of closures on the four components of the hospital quality composite from Hospital Compare. We follow Doyle et al. (2019) in constructing the four measures: process measures, patient survey measures, 30-day risk-adjusted mortality rates and 30-day risk-adjusted readmission rates. More detail on the Hospital Compare measures can be found in Section A.2.1. The remaining plots show effects of closures on the components of the three infant/maternal morbidity composites. "U" represents measures from the unrevised birth certificates and "R" represents measures from the revised birth certificates. Section A.3 provides more details on these measures, and Table A8 details the years and the number of states for which each of these variables is available.

Figure A4: Effect of Closures on C-Section Delivery by Predicted C-Section Risk



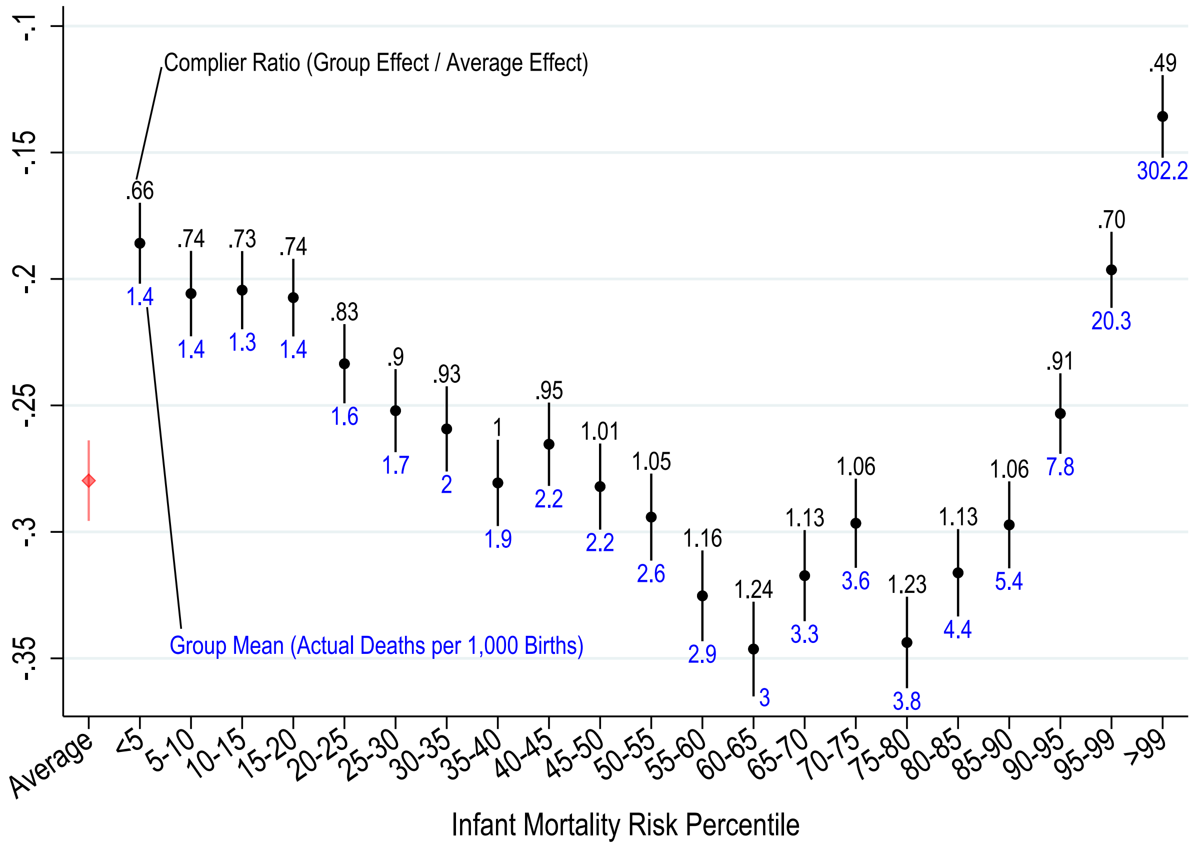
Notes: Means represent the (mean) predicted C-section risk for each quartile. C-section risk is calculated for each birth as the predicted value from an individual-level logistic regression of C-section delivery on the following risk factors (all indicator variables): 5-year maternal age bands, birth order (up to 5), singleton, breech, eclampsia, chronic hypertension, pregnancy hypertension, diabetes, and previous C-section delivery. Previous C-section delivery, which is a critical predictor of C-section risk, could not be calculated for state-years using the unrevised birth certificates after 2009. As such, those state-years are omitted in these estimates (approximately 2.8% of the sample).

Figure A5: Nonlinear (Quadratic) Marginal Effects of Closures



Notes: Each plot represents a separate regression in which the Closed indicator in Eq. (1) is replaced with a quadratic in the crow-flies distance to the nearest OB unit. Dashed lines represent the predicted difference in the outcome between having an operational OB unit *X* miles from a mother's county of residence and having one in her county of residence (i.e., zero miles). Shaded regions represent 95% confidence intervals. Note that the quadratic specification exploits broader variation compared to the closure indicator. Specifically, distance to the nearest OB unit can also arise due to openings (of which there are a small number) or due to closures/openings in nearby counties (if a mother's own county lacks a OB unit). The infant and maternal morbidity outcomes are composite measures, where "U" represents measures from the unrevised birth certificates and "R" represents measures from the revised birth certificates. For reference, the following values represent percentiles in distance to the nearest OB unit in the first year following a county's closure: 30.7 miles (25th), 37.1 miles (50th), 46.1 miles (75th), 59.8 miles (90th), 67.9 (95th).

Figure A6: Effect of Closures on Birth in County of Residence (First Stage) by Predicted Mortality Risk



Notes: The leftmost estimate (in red) represents the average effect across risk groups. Remaining estimates correspond to percentiles of infant mortality risk. The numbers above each point (in black) represent the complier ratio: the subgroup estimate divided by the average effect. The numbers below each point (in blue) represent the actual (not predicted) number of deaths per 1,000 live births for each risk group. Infant mortality risk is calculated using predicted values from an individual-level logistic regression of infant mortality on: gestation week indicators, 5-year age bands, birth order indicators, singleton, breech, eclampsia, chronic hypertension, pregnancy hypertension, and diabetes. The pseudo-R^2 from this regression is 0.32 and most of the predictive power is generated through the gestation week indicators.

Appendix Tables

Table A1: Effects of Closures using AHA-based Coding of Closures (1995-2016)

	Panel A: Birth Location, Prenatal Visits and Birthweight					
	Birth in Cnty. of Residence	Birth in Hospital	Prenatal Visits	Birthweight	Low Bir. Wt.	V. Low Bir. Wt.
Closed	-0.258***	-0.000958	-0.153***	-2.484	0.0000935	0.0000172
	(0.00884)	(0.000945)	(0.0473)	(2.154)	(0.000860)	(0.000314)
N	59,840	59,840	59,839	59,840	59,840	59,840
	Panel B: Gestation and Induction					
	Weeks Gestation	Premature (<37 Weeks)	Gestation 37-39 Weeks	Induced at 37-39 Weeks	Induced Ever	
Closed	-0.0328***	0.00156	0.00847***	0.0159***	0.0123***	
	(0.0112)	(0.00110)	(0.00237)	(0.00365)	(0.00245)	
N	59,840	59,840	59,840	59,840	59,840	
	Panel D: Maternal and Infant Health Outcomes					
	Cesarean	Low APGAR	Infant Morbid. (Unrevised)	Infant Mortality Rate	Maternal Morbid. (Unrevised)	Maternal Morbid. (Revised)
Closed	-0.0117***	-0.00228	0.0140	0.111	0.0145	-0.0399
	(0.00237)	(0.00164)	(0.0381)	(0.227)	(0.0288)	(0.0361)
N	59,840	58,602	28,397	59,840	31,439	17,220
	Panel C: Birth Environment					
	HC Composite in Birth Cnty.	NICU in Birth Cnty.				
Closed	0.0614***	0.0389***				
	(0.0233)	(0.0103)				
N	16,016	59,774				

Note: Estimates come from the two-way fixed effects (TWFE) specifications displayed in Figures 2–4, but the treatment (closures) is constructed using AHA data (as opposed to NVSS data as in the main specification). The AHA sample runs from 1995 (the first year addresses were available) through 2016.

Table A2: Mean Outcomes and County Characteristics

	All Counties	Closure Counties	Non-Closure Counties Unweighted	Non-Closure Counties P-Weighted
Panel A: County Characteristics				
Fertility Rate	66.91	67.83	66.68	67.46
Fertility Rate Growth Rate	0.0081	0.0095	0.0077	0.0160
Population	91,518	22,278	109,321	22,206
Population Growth Rate	0.0050	0.0015	0.0058	0.0027
Empl./Pop.	0.51	0.49	0.51	0.49
Percent Urban	0.41	0.29	0.44	0.28
Female 15-44 Pop. Share	0.38	0.36	0.38	0.36
Panel B: Birth Location, Prenatal Care and Outcomes Determined Prior to Birth				
Occurrence in Cnty. of Res.	0.39	0.19	0.44	0.28
Occurrence in Hospital	0.985	0.985	0.985	0.985
Number of Prenatal Visits	11.20	11.05	11.24	11.07
Birthweight	3,300	3,300	3,299	3,300
Low Birthweight (<2500g)	0.076	0.076	0.076	0.075
V. Low Birthweight (<1500g)	0.013	0.013	0.013	0.013
Weeks Gestation	38.76	38.74	38.76	38.75
Premature (<37 Weeks)	0.12	0.12	0.12	0.12
Gestation 37-39 Weeks	0.51	0.51	0.51	0.51
Induced	0.24	0.25	0.24	0.25
Induced at 37-39 Weeks	0.12	0.13	0.12	0.13
Panel C: Birth Environment in County of Occurrence				
Hospital Quality in Bir. Cnty.	0.056	0.051	0.057	0.035
NICU in Bir. Cnty.	0.42	0.40	0.42	0.34
Panel D: Health Outcomes				
Cesarean Delivery	0.28	0.28	0.28	0.28
Low Apgar (<8)	0.04	0.04	0.04	0.04
Infant Mortality Rate	7.25	7.47	7.19	7.24
Infant Morbidity (U)	0.0029	-0.0078	0.0056	-0.0148
Maternal Morbidity (U)	-0.0001	-0.0272	0.0069	-0.0133
Maternal Morbidity (R)	0.000	-0.0122	0.0031	0.0163
Number of Counties	2,958	605	2,353	2,353

The fourth column ("Non-Closure P-Weighted") weights by the propensity to experience a closure. Weighting forces similarity between treated and untreated counties. It ensures, for example, that the comparison group for the largely rural treated counties is also largely rural. The exact process of calculating the weights is described in Section A.4.2. "U" represents measures from the unrevised birth certificates and "R" represents measures from the revised birth certificates.

Table A3: Effects of Closures with Alternative Treatment Definition & Keeping Counties with Reopenings

	Panel A: Birth Location, Prenatal Visits and Birthweight					
	Birth in Cnty. of Residence	Birth in Hospital	Prenatal Visits	Birthweight	Low Bir. Wt.	V. Low Bir. Wt.
No OB Unit	-0.307***	-0.00178***	-0.151***	-0.634	-0.000781	-0.00000572
	(0.00652)	(0.000688)	(0.0347)	(1.484)	(0.000537)	(0.000183)
N	91,383	91,383	91,383	91,383	91,383	91,383
	Panel B: Gestation and Induction					
	Weeks Gestation	Premature (<37 Weeks)	Gestation 37-39 Weeks	Induced at 37-39 Weeks	Induced Ever	
No OB Unit	-0.0335***	0.000333	0.00939***	0.0135***	0.0103***	
	(0.00738)	(0.000735)	(0.00158)	(0.00247)	(0.00167)	
N	91,383	91,383	91,383	91,310	91,310	
	Panel D: Maternal and Infant Health Outcomes					
	Cesarean	Low APGAR	Infant Morbid. (Unrevised)	Infant Mortality Rate	Maternal Morbid. (Unrevised)	Maternal Morbid. (Revised)
No OB Unit	-0.0102***	-0.00377***	0.0330	0.0522	-0.00419	-0.0701***
	(0.00167)	(0.00112)	(0.0201)	(0.141)	(0.0184)	(0.0271)
N	91,589	89,473	46,444	91,667	51,842	27,590
	Panel C: Birth Environment					
	HC Composite in Birth Cnty.	NICU in Birth Cnty.				
No OB Unit	0.0984***	0.0502***				
	(0.0214)	(0.00820)				
N	24,690	64,988				

Note: In the main specification, the treatment (“Closed”) is an indicator equal to one in all years following closures (treatment never switches off, as assumed in a standard staggered DD design), and counties in which OB units reopen are dropped from the sample. In this alternative specification, the treatment (“No OB Unit") is equal to one in all counties and years in which there is no operational OB unit and we include all counties including those that experience a reopening. As such, this specification allows treatment to switch on and off and thus uses more variation (including openings). This type of treatment variable, however, is not compatible with recent alternative DD estimators (de Chaisemartin and D’Haultfoeuille, 2020; Borusyak et al., 2021).

Table A4: Specification Checks: Birth Location, Prenatal Care, and Outcomes Determined Prior to Delivery Experience

	(1)	(2)	(3)	(4)	(5)
Birth in Cnty. of Residence	-0.302***	-0.301***	-0.300***	-0.301***	-0.298***
	(0.00719)	(0.00716)	(0.00714)	(0.00710)	(0.00691)
Birth in Hospital	-0.00262***	-0.00187**	-0.00207**	-0.00173**	-0.000933
	(0.000861)	(0.000874)	(0.000859)	(0.000813)	(0.000767)
Prenatal Visits	-0.163***	-0.163***	-0.164***	-0.177***	-0.193***
	(0.0404)	(0.0407)	(0.0407)	(0.0387)	(0.0375)
Birth Weight	-3.987**	-2.575	-2.176	-1.145	-2.122
	(1.737)	(1.784)	(1.778)	(1.759)	(1.783)
Low Birth Wt. (<2500g)	-0.0000617	-0.000192	-0.000123	-0.000162	0.0000739
	(0.000620)	(0.000645)	(0.000640)	(0.000635)	(0.000661)
Very Low Birth Wt. (<1500g)	0.000310	0.000258	0.000228	0.000158	0.000231
	(0.000216)	(0.000222)	(0.000222)	(0.000223)	(0.000233)
Weeks Gestation	-0.0569***	-0.0473***	-0.0447***	-0.0340***	-0.0312***
	(0.00842)	(0.00860)	(0.00855)	(0.00834)	(0.00848)
Premature (<37 Weeks)	0.00202**	0.00146*	0.00150*	0.000601	0.000427
	(0.000842)	(0.000866)	(0.000871)	(0.000870)	(0.000907)
Gestation 37-39 Weeks	0.0131***	0.0112***	0.0103***	0.00878***	0.00804***
	(0.00185)	(0.00188)	(0.00184)	(0.00176)	(0.00181)
Induced at 37-39 Weeks	0.0157***	0.0137***	0.0129***	0.00983***	0.00887***
	(0.00194)	(0.00195)	(0.00190)	(0.00168)	(0.00170)
Induced	0.0206***	0.0185***	0.0175***	0.0135***	0.0125***
	(0.00285)	(0.00286)	(0.00281)	(0.00243)	(0.00245)
N	87,053	87,053	87,048	87,048	86,588
Sample Years			1989-2019		
County FE	X	X	X	X	X
Year FE	X	-	-	-	-
Urban-Year FE	-	X	X	X	X
County Controls	-	-	X	X	X
State-Year FE	-	-	-	X	X
P-Score Weight	-	-	-	-	X

Notes: Each row represents a different outcome and each column represents a different specification. For reference, Column 3 is the baseline specification. ***, **, * indicate significance at the 1%, 5% levels, and 10% levels.

Table A5: Specification Checks: Infant and Maternal Health Outcomes

	(1)	(2)	(3)	(4)	(5)
Cesarean	-0.0107***	-0.0108***	-0.0106***	-0.0108***	-0.0114***
	(0.00202)	(0.00205)	(0.00204)	(0.00193)	(0.00187)
N	86,980	86,980	86,975	86,975	86,515
Sample Years			1989-2019		
Low Apgar (<8)	-0.000652	-0.00247**	-0.00217*	-0.00224*	-0.00229*
	(0.00125)	(0.00125)	(0.00124)	(0.00119)	(0.00121)
N	85,049	85,049	85,044	85,044	84,584
Sample Years			1989-2019		
Infant Mortality Rate	0.179	0.161	0.145	0.103	0.128
	(0.166)	(0.175)	(0.171)	(0.168)	(0.178)
N	87,053	87,053	87,048	87,048	86,588
Sample Years			1989-2019		
Neonatal Mortality Rate	0.0782	0.0395	0.0173	0.00418	0.0183
	(0.138)	(0.144)	(0.141)	(0.141)	(0.147)
N	87,053	87,053	87,048	87,048	86,588
Sample Years			1989-2019		
Infant Composite (1989-2006)	0.0484*	0.0487*	0.0515*	0.0693***	0.0658***
	(0.0266)	(0.0268)	(0.0273)	(0.0266)	(0.0247)
N	44,219	44,219	44,214	44,214	43,991
Sample Years			1989-2006		
Maternal Composite (1989-2006)	-0.00358	0.00533	0.00621	0.0104	0.0162
	(0.0222)	(0.0228)	(0.0229)	(0.0230)	(0.0211)
N	49,262	49,262	49,257	49,257	48,996
Sample Years			1989-2006		
Maternal Composite (2009-2019)	-0.102***	-0.105***	-0.103***	-0.0995***	-0.0949***
	(0.0299)	(0.0302)	(0.0305)	(0.0302)	(0.0308)
N	26,170	26,170	26,170	26,170	26,034
Sample Years			2009-2019		
County FE	X	X	X	X	X
Year FE	X	-	-	-	-
Urban-Year FE	-	X	X	X	X
County Controls	-	-	X	X	X
State-Year FE	-	-	-	X	X
P-Score Weight	-	-	-	-	X

Notes: Each panel represents a different outcome and each column represents a different specification. For reference, Column 3 is the baseline specification. ***, **, * indicate significance at the 1%, 5% levels, and 10% levels.

Table A6: Specification Checks: Quality Measures

	(1)	(2)	(3)	(4)	(5)
HC Composite in Birth Cnty.	0.103***	0.102***	0.101***	0.101***	0.0984***
	(0.0230)	(0.0229)	(0.0228)	(0.0225)	(0.0225)
N	23,710	23,710	23,710	23,710	23,560
Sample Years			2010-2019		
NICU in Birth Cnty.	0.0382***	0.0436***	0.0423***	0.0430***	0.0410***
	(0.00986)	(0.00952)	(0.00947)	(0.00879)	(0.00830)
N	61,710	61,710	61,710	61,710	61,380
Sample Years			1995-2016		
County FE	X	X	X	X	X
Year FE	X	-	-	-	-
Urban-Year FE	-	X	X	X	X
County Controls	-	-	X	X	X
State-Year FE	-	-	-	X	X
P-Score Weight	-	-	-	-	X

Notes: Each row represents a different outcome and each column represents a different specification. For reference, Column 3 is the baseline specification. ***, **, * indicate significance at the 1%, 5% levels, and 10% levels.

Table A7: Specification Checks: Quality Measures

	Gestation 37-38 Weeks	Gestation 39 Weeks	Gestation 40+ Weeks	Induction at 37-38 Weeks	Induction at 39 Weeks	Induction at 40+ Weeks
Closed	0.00351*	0.00684***	-0.0118***	0.00470***	0.00823***	0.00311*
	(0.00146)	(0.00137)	(0.00199)	(0.000955)	(0.00125)	(0.00129)
N	87,048	87,048	87,048	86,977	86,977	86,977

Notes: . ***, **, * indicate significance at the 1%, 5% levels, and 10% levels.

Table A8: Number of States Reporting Maternal and Infant Health Measures

	Infant Comp. (1989-2006)				Maternal Comp. (1989-2006)			Maternal Comp. (2009-2019)				
	Meconium	Injury	Seizure	Vent.	Fever	Bleeding	Seizure	Transfus.	Lacerat.	Rupture	Hyster.	ICU
1989-2002	47	45	47	47	47	47	47	0	0	0	0	0
2003	47	43	47	47	45	45	45	0	0	0	0	0
2004	47	46	47	47	46	46	46	0	0	0	0	0
2005	47	47	47	47	47	47	47	0	0	0	0	0
2006	47	45	47	47	45	45	45	0	0	0	0	0
2007	47	0	47	47	0	0	0	0	0	0	0	0
2008	47	0	47	47	0	0	0	0	0	0	0	0
2009	47	0	47	47	0	0	0	19	19	19	19	19
2010	47	0	47	47	0	0	0	24	24	24	24	24
2011	47	0	47	47	0	0	0	29	29	29	29	29
2012	47	0	47	47	0	0	0	31	31	31	31	31
2013	47	0	47	47	0	0	0	35	35	35	35	35
2014	0	0	47	47	0	0	0	43	43	43	43	43
2015	0	0	47	47	0	0	0	44	44	44	44	44
2016-2019	0	0	47	47	0	0	0	47	47	47	47	47

Note: The maximum number of states is 47 because we drop states outside the contiguous US (HI and AK), and we drop Virginia because counties are defined differently in Virginia ("townships" instead of counties) and their boundaries have changed significantly over time. "Meconium" refers to meconium staining; "Vent." refers to infant use of ventilator; "Transfus." refers to maternal transfusion; "Lacerat." refers to 3rd or 4th degree perineal lacerations; "Rupture" refers to ruptured uterus; "Hyster." refers to unplanned hysterectomy; "ICU" refers to maternal admission to the ICU.

Table A9: Example of Identifying a Closure

Year	Number of Hospital-Based Births Occurring in County X	Number of Hospital-Based Births Occurring in County Y	Closed County X	Closed County Y
1995	142	142	0	0
1996	153	153	0	0
1997	114	114	0	0
1998	125	125	0	0
1999	107	107	0	0
2000	118	118	0	0
2001	55	7	1	1
2002	4	4	1	1
2003	1	1	1	1
2004	0	0	1	1
2005	0	0	1	1
2006	2	2	1	1
2007	1	1	1	1

Notes: This representative example uses fabricated data due to confidentiality. Both County X and County Y are coded as open 1995-2000 and closed 2001-2007. The rule used to identify closures, which is outlined in Section A.1.2, deals well with County X. In County X, hospital-based births declined by at least 75% between 2001 and 2002, there were more than 6 births in 2001 and less than 6 births in 2002 (there were 55 births in 2001 and only 4 in 2002). As such, in 2001 County X meets the rule for a closure. While the closure rule identifies most closures, there are a few cases that require manual coding. For instance, in 2001 there were 7 births in County Y and in 2002 there were only 4. While there were more than 6 births in 2001 and fewer than 6 births in 2002, there was not at least a 75% reduction between 2001 and 2002. Consequently, the rule codes County Y as open in 2001 when in fact it was clearly closed. There were 100+ births 1995-2000, and virtually no births starting in 2001. The most common reason for needing manual coding of closures is due to closures occurring early in the year. When this occurs, births dramatically decline in this partially closed year but they do not necessarily immediately drop to near zero.

Table A10: Closure Logistic Regression Estimates

Fertility Rate	-0.00314
	(0.00245)
Emp./Pop. Ratio	-0.512
	(0.269)
Earnings Per-Capita	-0.00353
	(0.0153)
Transfers Per-Capita	0.0993
	(0.0805)
Female Pop. Share 15-19	2.484
	(3.808)
Female Pop. Share 20-24	-11.36***
	(3.282)
Female Pop. Share 25-29	4.033
	(5.350)
Female Pop. Share 30-34	-6.353
	(5.918)
Female Pop. Share 35-39	-7.766
	(5.795)
Female Pop. Share 40-44	-4.306
	(5.502)
Total Pop.	-0.00000892***
	(0.00000164)
Pop. Density	0.0000677
	(0.000446)
Percent urban	0.00220
	(0.00139)
N	2,947
Pseudo R^2	0.106

Notes: ***, **, * indicate significance at the 1%, 5% levels, and 10% levels. Estimates are from a cross-sectional logistic regression where the outcome is an indicator for a county ever experiencing a closure. Regressors represent county characteristics in the first year of the sample (1989).

Appendix: Data and Econometric Approach

A.1 Data Appendix

A.1.1 Identifying Closures in the AHA Data

We use data from the AHA Annual Surveys for 1995-2016 to identify maternity ward closures at the hospital (address) level. While the AHA data are available for prior years as well, 1995 was the first year in which addresses were reported. There is no single variable in the AHA data that measures the presence of an operational obstetric unit (which could then be used to identify closures), instead we develop an algorithm to detect closures. The algorithm is based on three variables: the number of obstetric beds, the number of bassinets, and the number of births. This algorithm is necessary not only because there is no single variable measuring operational obstetric units, but also due to non-response in some of the measures (e.g., 17% of observations on obstetric beds are missing). Furthermore, the algorithm alleviates concerns about inaccurate responses, since the algorithm relies on agreement between multiple variables in the data. Let OB_{Open} be an indicator for the presence of an operational OB unit; the algorithm is defined as below:[16]

1. Set $OB_{Open} = 0$ if the hospital reports zero obstetric beds, zero bassinets, and < 10 births.
2. Set $OB_{Open} = 1$ if the hospital reports > 0 obstetric bed, > 0 bassinets, and > 10 births.
3. If OB_{Open} is still not defined, set $OB_{Open} = 0$ if the hospital reports < 5 births.
4. If OB_{Open} is still not defined, set $OB_{Open} = 1$ if the hospital reports > 25 births.
5. If OB_{Open} is still not defined, set $OB_{Open} = 0$ if the hospital reports zero bassinets.
6. If OB_{Open} is still not defined, set $OB_{Open} = 1$ if the hospital reports > 0 bassinets.

With information on the presence of an operational obstetric unit for each hospital, closures (i.e., the treatment variable) are defined as events in which OB_{Open} changes from 1 to 0. While our primary method of inferring closures is based on the NVSS data, we report results for all the main outcomes using the AHA-based method in Table A1. The results are qualitatively similar across all outcomes.

In addition to using the AHA data as an alternative method of identifying OB unit closures, we also use the data for information on hospital characteristics. Specifically, we use AHA data to identify the presence of neonatal intensive care units (NICUs) in each county. We use this information in our analysis of hospital quality and resources, and more details are provided on this aspect of the data in Section A.2.2.

[16] This algorithm classifies 100% of hospitals as either having an operational OB unit or not.

A.1.2 Identifying Closures in the NVSS Data

While the AHA data has advantages (i.e., hospital addresses and information on hospital characteristics), the survey nature of the data may induce substantial measurement error. Furthermore in the AHA data, hospitals within the same system but in different locations are sometimes coded with the same address, limiting our ability to precisely identify local closures in this data. A more reliable method of identifying hospital-level closures would be to use hospital-level administrative records of births and infer a closure when there is a sudden drop in the number of births. While these data do not exist for the entire US, the NVSS data do cover the entirety of the United States and include information on both county of residence and county of occurrence. This allows us to identify whether there are any operational OB units in a given county, which is our main treatment variable.[17]

To identify OB unit closures in the NVSS data, we look for events in which the number of *hospital-based* births *occurring* in a county drops to near zero.[18] To achieve this, we use a simple rule to identify closures: for a particular county, we identify year y as the year of a closure if the number of hospital-based births declined by at least 75% between year y and year $y+1$, where the number of births in year y was at least six, and the number of births in year $y+1$ was less than six. We use a similar symmetric rule to identify openings: for a particular county, we identify year y as the year of an opening if the number of hospital-based births increased by at least 300% between year y and year $y+1$, where the number of births in year y was less than six, and the number of births in year $y+1$ was more than six. While these simple rules identify most closures, there were a number of cases that were not identified by these rules, and we code those manually. In total, we identify 640 counties with either an opening or closure, and we manually adjusted closure or opening dates for 151 of these.

Table A9 provides an example (with fabricated data, for confidentiality) of our method for identify OB unit closures for two counties. In both cases, we code the year of closure as 2001. For county X, this is identified by the rule, but for county Y it is not and, thus, requires manual coding. Specifically, in County X, hospital-based births declined by at least 75% between 2001 and 2002, and there were more than 6 births in 2001 and less than 6 births in 2002 (there were 55 births in 2001 and only 4 in 2002). As such, in 2001 County X meets the rule for a closure and is coded as closed. On the other hand, in County Y there were 7 births in 2001 and 4 in 2002. While there were more than 6 births in 2001 and fewer than 6 births in 2002, there was not at least a 75% reduction between 2001 and 2002. Consequently, the rule codes County Y as open in 2001 when in fact it

[17] Notably, we cannot use these data with some alternative definitions of the treatment. For example, we cannot identify the number of operational OB units in a county.

[18] To be clear, in the NVSS data we observe each mother's county of residence and the county of birth occurrence; the algorithm utilizes only the county of birth occurrence. The data also contain information on whether each birth takes place in a hospital or other setting, and the algorithm utilizes only births in hospitals.

is clearly closed. There were 100+ births 1995-2000, and virtually no births starting in 2001. The most common reason for needing manual coding of closures is due to closures occurring early in the year. When this occurs, births dramatically decline in this partially closed year but they do not necessarily immediately drop to near zero.

A.2 Measures of Hospital Quality & Resources

Our hospital quality metrics are grouped into three categories: (1) measures based on Centers for Medicare and Medicaid (CMS) Hospital Compare, (2) risk-adjusted infant mortality, and (3) the presence of a NICU.

A.2.1 Hospital Compare Measures

Quality metrics from Hospital Compare are publicly-available, hospital-level measures that have been widely used and scrutinized (e.g., Chandra et al. (2016)). In an analysis evaluating these metrics, Doyle et al. (2019) find that patients pseudo-randomly assigned to hospitals with higher hospital quality metrics do indeed achieve better outcomes, suggesting these are useful measures of hospital quality.

Hospital Compare provides several quality measures, and we generally follow Doyle et al. (2019) in constructing the following four measures at the hospital level (exceptions described below): process measures, patient survey measures, 30-day risk-adjusted mortality rates and 30-day risk-adjusted readmission rates. While we provide the necessary information here, please see Doyle et al. (2019) for a more detailed description of these data.

Process measures are scores based on the extent to which hospitals implement specific best-practices. For example, one score is based on whether heart attack (AMI) patients were given Aspirin at discharge. We follow Doyle et al. (2019) and define our process measure as the average of seven scores based on hospital practices for heart attack, heart failure, pneumonia, and surgery:

1. Heart failure patients given ACE inhibitor or ARB for left ventricular systolic dysfunction.
2. Heart attack (AMI) patients given Aspirin at discharge.
3. Heart failure patients given assessment of left ventricular function.
4. Heart failure patients given discharge instructions.
5. Pneumonia patients given the most appropriate initial antibiotic.
6. Surgery patients who received preventative antibiotic(s) one hour before incision.
7. Surgery patients whose preventative antibiotic(s) are stopped within 24 hours after surgery.

Patient Survey measures provided in Hospital Compare are derived from the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey. The survey covers a range of aspects regarding the patient's experience at the hospital. Again, we follow Doyle et al. (2019) and define our survey measure as the average of ten individual survey scores:

1. Doctors always communicated well.
2. Nurses always communicated well.
3. Pain was well controlled.
4. Patients always received help as soon as they wanted.
5. Patients gave an overall rating of 9 or 10 (high).
6. Room was always clean.
7. Room was always quiet at night.
8. Staff always explained.
9. Yes, patients would definitely recommend the hospital.
10. Yes, staff did give patients this information.

The two outcome-based measures are risk-adjusted rates of mortality and readmission within 30 days of discharge (the measures are transformed so that higher values represent higher quality). For these measures, we depart from Doyle et al. (2019) in one respect: while they use mortality/readmission rates for AMI, heart failure and pneumonia, we use mortality/readmission rates only for heart failure and pneumonia. The reason is that mortality/readmission rates for AMI are missing for a substantial number of hospitals. For example, when aggregated to the county level, we have valid observations from only 1,161 counties for the measure that includes AMI compared to 1,672 counties for the measure that excludes AMI. Since our analysis focuses on (often small) rural counties and hospitals, it is extremely important to maintain as broad of coverage as possible.

Hospital Compare data has been released in numerous waves (with multiple per year in many years), beginning in March 2010. Each release of the data represents data measured in prior years, where the years represented depends on the measure. For example, the March 2010 release represented process and survey measures from July 2008-June 2009, and mortality and readmission measures from 2005-2008. Following Doyle et al. (2019), we maintain these lags and assign each hospital its average measure across a number of waves. Specifically, we average across all five waves released in 2010. As such, our quality metrics are time-invariant (and we limit our analysis sample to 2010-2019). We use these time-invariant measures for three reasons. First, by only using measures from a period prior to our analysis period, this ensures the quality metrics are not endogenous to OB unit closures. Second, specific measures have been phased out over time; for example, when aggregated to the county-year level, we observe process measures for 1,551 counties in the 2010 waves, 979 in the 2013 waves, and this measures is gone completely by 2016. Third, the process measures have become less meaningful over time; Doyle et al. (2019) show the process measures became extremely compressed at the top of the distribution by 2015, as hospitals were able to respond to these publicly-reported metrics by updating their processes.

After constructing these hospital-level measures, we then aggregate to the county level to match our level of analysis, weighting by the number of beds in each hospital. As such, our measures represent the bed-weighted average hospital quality for a given county. We derive information on the location and bed count for each hospital from the Medicare Provider of Service files. Finally, in order to construct an overall, county-level proxy for quality, we create a composite of the four measures. The composite is created by standardizing each measure at the county-level (Mean=0, SD=1), then taking a simple average of the z-scores. We use this composite for three reasons: (1) we are not necessarily interested in the specific measures of hospital quality, but rather a general proxy for quality, (2) by constructing a composite, we can potentially increase the power of our estimates , and (3) to simplify exposition.

A.2.2 NICU

We use the presence of a neonatal intensive care unit (NICU) in the county of birth occurrence as a measure of obstetric-specific hospital resources (rather than quality, per se). This information is derived from the AHA Annual Surveys for 1995-2016. In this hospital-level survey data, hospital-years are defined as having an operational NICU if there is any NICU beds. Because this is survey data, 17.3% of hospital-years have missing information on the number of NICU beds. We code NICU status and impute missing values using the following algorithm:

1. For hospital-years with non-missing data, assign NICU=1 for those with at least one bed, and NICU=0 for those with none.
2. For hospital-years with missing data, assign NICU=1 if NICU=1 for the hospital in every other year.
3. For hospital-years with missing data, assign NICU=0 if NICU=0 for the hospital in every other year.
4. For hospital-years with missing data, assign NICU=0 if the hospital has no non-missing values for any year.
5. For hospital-years with missing data, assign NICU equal to the hospital's most recent non-missing value.
6. For hospital-years with missing data, assign NICU equal to the hospital's closest future non-missing value.

A.3 Infant and Maternal Health Composite Measures

We analyze a range of outcomes that measure maternal and infant morbidity, but our primary measures are composites constructed from multiple outcomes. In constructing these, we are limited to the states and years in which all components of each composite are observed. With the state-level rollout of the revised birth certificate in 2003, several variables were either phased in or out. Table A8 describes the number of states in which each component of the composite measure is available for each year of the sample. This table shows that several components that were widely available prior to the revision phase out completely in 2006 (hence the reason two of the composite measures are only defined through that year). Likewise, several maternal morbidity measures were only phased in beginning in 2009 (beginning with 19 states reporting). Also note that some components of the infant composite are available for the entire sample; specifically, while the infant composite measure is only defined for any state between 1989-2006, two components of the measure (infant seizures and use of ventilator) are observed for the entire sample.

A.4 Details of the Econometric Approach

A.4.1 Two-Way Fixed Effects & Negative Weights

A recent literature has shown that applying TWFE approaches to DD designs can lead to biased estimates (e.g., Goodman-Bacon (2021); Borusyak et al. (2021); de Chaisemartin and D'Haultfoeuille (2020)). Simplifying the problem, this issue is largely due to the fact that the TWFE approach is a weighted average of average treatment effects on the treated (ATTs) from many two-by-two DD comparisons, where some of the weights can be negative when treatment effects are heterogeneous. Negative weights arise from poor comparisons such as those between treated units and previously-treated units, whereas comparisons between treated units and never-treated units are arguably more clean. This negative weighting issue is particularly problematic in settings with few or zero never-treated units, since the number of "clean" comparisons is limited in those settings. Fortunately, in our setting, most counties never experience a closure and thus are never treated. This means the potential for the negative weighting issue to bias our TWFE estimates is limited. We confirm this intuition by using the de Chaisemartin and D'Haultfoeuille (2020) procedure to test for the presence of negative weights. Specifically, we implement this approach for the most parsimonious TWFE specification (i.e., county and time fixed effects with no time-varying covariates) and using the first-stage outcome (i.e., the share of mothers giving birth in their county of residence). We find that the average estimate is a weighted sum of 10,531 ATTs, where 588 (5.6%) of those receive negative weight. While that is a small but non-zero proportion of ATTs receiving negative weight, their importance is close to zero: the negative weights sum to -0.0058 (all weights sum to 1).

While we do not expect the TWFE estimates to be substantially biased in our setting, we present estimates from two alternative estimators that are robust to the negative weighting issue. Results from the de Chaisemartin and D'Haultfoeuille (2020) estimator are presented alongside the main results in Figures 2–5.

A.4.2 Alternative Specifications

While our main empirical specification is described in Eq. (1), we also include a range of alternatives and present the results for all of the main outcomes in Tables A4 to A6. The specifications in each of the five columns of these tables are described below.

1. A parsimonious TWFE specification, including only county and year fixed effects.

2. The baseline specification, but excluding time-varying covariates.

3. The baseline specification.

4. The baseline specification, plus state-by-year fixed effects. These control for any factors specific to a state (but common to all counties within the state) that vary over time, such as a state's decision to expand Medicaid following passage of the Affordable Care Act.

5. The specification in column 4, but weighting untreated counties by their treatment propensity. We estimate this specification because one might be concerned that counties experiencing closures might not be comparable to counties that do not. This specification forces comparability between treatment and comparison counties. To implement this, we predict the probability of ever experiencing a closure in a cross-sectional county-level logistic regression based on a set of county-level characteristics observed in the first year of the sample, 1989. We then weight the untreated counties by $\frac{\hat{p}}{(1-\hat{p})}$, where $\hat{p}$ is the predicted probability of experiencing a closure from the logit (treated observations receive weight equal to one). This effectively gives more weight to rural counties and essentially zero weight to dense and highly populated urban counties. The estimates from the predictive regression are shown in Table A10.

A.4.3 Event Study Specification

$$\mathrm{Y}_{cy} = \sum_{j=-8}^{-2} \beta_j \mathrm{Closed}_{cyj} + \sum_{j=0}^{8} \beta_j \mathrm{Closed}_{cyj} + \gamma \mathrm{X}_{cy} + \delta_c + \delta_{uy} + \varepsilon_{cy} \tag{2}$$

The event study version of our TWFE specification is described in the equation above. Specifically, this specification is the same as Eq. (1) except that we have replaced the single post-treatment indicator ($Closed_{cy}$) with a set of 16 indicators for time relative to treatment, Closed_{cyj}. The indicator for one year prior to treatment is omitted as the reference group. The two end points ($j = -8$ and $j = 8$) represent eight *or more* years prior to treatment and eight *or more* years post-treatment and, as such, the specification is fully saturated. Because the end points are not comparable with the other estimates, the end points are omitted from the figures displaying the results. Some outcomes are only observed for a subset of the sample (e.g., the Hospital Compare quality metrics). For outcomes with a significantly limited sample, we include 10 indicators for time relative to treatment (i.e., $j = -5$ to $j = 5$, omitting $j = -1$) and report estimates for four years pre- and post-treatment.

A.4.4 Sample Restrictions for C-Section Mechanism Analysis

This section refers to the estimates presented in Figure 4. Like all of the main analyses, closure counties that experienced an opening at any point in the sample are omitted (117 counties omitted, leaving 488 in the analysis sample). This analysis requires restricting the sample in three additional ways.

1. The first three years (1989-1992) of the overall sample are dropped to account for the fact that the outcome in Figure 4A and the C-section gaps in Figure 4B utilize 3-year lags in C-section rates.

2. The sample is limited to state-years in which it is possible to calculate risk-adjusted C-section rates. Previous C-section delivery, which is a critical predictor of C-section risk, could not be calculated for state-years using the unrevised birth certificates after 2009. As such, those state-years are omitted in these estimates (approximately 2.8% of the sample is omitted).

3. The sample of counties experiencing a closure is limited to those that ever offered C-section delivery. 68 closure counties (14% of the 488 in the main analysis sample) recorded zero C-section deliveries in at least one of the three years prior to closure. The analysis does not have the same interpretation for those counties since all women in need of C-section delivery would have traveled outside of the county to give birth in the years prior to closure.