NBER WORKING PAPER SERIES

HEALTH CARE CENTRALIZATION:
THE HEALTH IMPACTS OF OBSTETRIC UNIT CLOSURES IN THE US

Stefanie J. Fischer
Heather Royer
Corey D. White

Working Paper 30141
http://www.nber.org/papers/w30141

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2022, Revised March 2023

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), SOLE Annual Meeting (2020), AEA Annual Meetings (2021), Montana State University Applied Micro Workshop, ASHEcon Annual Meetings (2022), NBER Health Economics Fall Program (2022), University of Melbourne (2022), Monash University (2022), University of California Santa Barbara (2022), University of Virginia (2022), Clemson University (2022), University of Nebraska (2022), All-California Labor Economics Conference (2022), University of Illinois-Urbana Champaign (2022), University of Southern California Center for Economic and Social Research (2022), and American University (2022). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Health Care Centralization: The Health Impacts of Obstetric Unit Closures in the US
Stefanie J. Fischer, Heather Royer, and Corey D. White
NBER Working Paper No. 30141
June 2022, Revised March 2023
JEL No. I18,I38,J08,J13,J18

ABSTRACT

Over the last few decades, health care services in the United States have become more geographically centralized. We study how the loss of hospital-based obstetric units in over 400 counties affect maternal and infant health via a difference-in-differences design. We find that closures lead mothers to experience a significant change in birth procedures 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.

Stefanie J. Fischer
Monash University
Department of Economics
900 Dandenong Rd, Building H
Caulfield East VIC 3145
Victoria
Australia
stefanie.fischer@monash.edu

Heather Royer
Department of Economics
University of California, Santa Barbara
2127 North Hall
Santa Barbara, CA 93106
and NBER
royer@econ.ucsb.edu

Corey D. White
Department of Economics
Monash University
Melbourne, VIC
Australia
corey.white@monash.edu

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 regionalization in 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 public 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 provider may offer better or worse services compared to 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 a 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 deliver in their county of residence (28 percentage point decrease), reduce the number of prenatal

[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.

care visits (0.17 fewer visits), and increase the probability of slightly earlier delivery due to the raised likelihood of scheduled induction (1.7 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 1.1 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 morbidity (maternal transfusions and 3rd/4th degree perineal lacerations decline by 0.19 and 0.28 percentage points, respectively).

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 (0.9 percentage point decrease), higher quality hospitals (0.1 standard deviation increase in our hospital quality index), and more obstetric-specific resources (4.2 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 are most likely to divert women to counties with lower relative 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 on obstetric unit closures, a common phenomenon across many developed countries. One challenge in this work is sample size. Narrowing in on one city, Philadelphia, between 1995 and 2005 where 9 out of 19 obstetric units closed, Lorch et al. (2013) estimate a 50% increase in neonatal mortality as a consequence of the closures, albeit their estimates are imprecise due to the limited sample. A broader geographic sample, which effectively increases the number of closures, is helpful for precision. Case in point, Kozhimannil et al. (2018) use an interrupted time series design with state fixed effects to study closures dispersed across the US. 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). In addition to the different empirical approach, expanded time frame, and new outcomes studied, 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, Avdic et al. (forthcoming) 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. 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 into 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). As hospital closures can perturb early life circumstances, this work also has relevance for the literature linking the perinatal environment to long-run health and human capital formation and adult outcomes (Almond et al., 2018).

2 Background on Closures

Hospital-based OB unit access has been in 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.[5] 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. The coincident trends documented in Figure 1A have been the subject of substantial attention from the media, think-tanks, and policy-makers, yet a causal relationship between these factors is not well-established.

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).[6] 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

[5] "Rural" counties are those classified as non-core or micropolitan in the 2013 NCHS urban/rural classification. While the closures we study primarily occur in rural counties, our empirical analysis is not restricted to only those counties.

[6] Carroll et al. (2022) find that Medicaid expansions over the last decade have not slowed the trend in obstetric unit closures.

(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).[7] 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.[8]

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 natality files provide information on both the infant and parents. We use the restricted-access version of the NVSS files which include information on the county of birth occurrence and the mother's county of residence (National Center for Health Statistics, 1989-2019a). These data also include information on whether the birth occurred in a hospital, the

[7]The national natality data do not include location of prenatal care visits, however, using administrative birth records data from Texas for the period 2000 to 2019, we calculate that 10-18 percent of mothers report obtaining prenatal care at a hospital facility depending on the year analyzed.

[8]It is plausible that effects of "crowding" are non-negative if increasing the number of births at receiving facilities results in economies of scale or opportunities for learning-by-doing.

number of prenatal visits, birth procedures (e.g., induction, C-section), and numerous measures of infant and maternal health. In our analysis, we examine several standard measures of infant health (gestational age, birthweight, Apgar scores) in addition to 12 other measures of infant and maternal morbidity (National Center for Health Statistics, 1989-2019b).

We construct three composite measures to summarize impacts on the 12 other infant and maternal morbidity outcomes. Many of the infant and maternal morbidity measures 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. We label our composite measures as either "Unrevised" (available 1989-2006) or "Revised" (available 2009-2019). Table A1 describes the number of states in which each component of each composite measure is available for each year of the sample. This table shows that several components that were widely available prior to the revision phased out completely in 2006. Beginning in 2009, five high-quality measures of maternal morbidity were phased in. The components of the unrevised infant morbidity composite include meconium staining, birth injury, infant seizure, and mechanical ventilation. The components of the unrevised maternal composite measure include maternal fever, excessive bleeding, and maternal seizure. The components of the revised maternal morbidity composite include maternal transfusion, 3rd/4th degree perineal lacerations, ruptured uterus, unplanned hysterectomy, and ICU admission.

The NVSS mortality files allow us to examine infant mortality rates. In most of our analysis, we use unlinked mortality files (i.e., deaths are not linked to births) since these files are available for the full 1989-2019 sample. In some analyses of mechanisms, we employ the linked birth-infant death files which are available for 1989-1991 and 1996-2017 (we note later where these data are used).[9]

While the features of the NVSS files described above make these data ideal for this analysis, there are limitations. One limitation is that the most granular geographic identifer is the county (county of residence and county of birth occurrence). Consequently, there is some measurement error in determining the nearest hospital once a closure has occurred. Ideally, one would observe the exact hospital of birth and the mother's exact address, but these data do not exist at a national scale. Another limitation is that the outcomes observed in the natality files, while quite broad, are limited. For example, we do not observe hospital diagnosis and procedure codes, which would

[9]Linked birth-infant death files are from National Center for Health Statistics (1989-1991 & 1996-2017).

allow us to measure a broader set of infant and maternal morbidity measures.[10] Conducting such an analysis at a national scale would require a nationwide census of hospital discharge data over a long time period.[11]

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 (American Hopstial Association, 1995-2016). 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 A2, 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 32 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). In other words, the NICU metric is an indicator

[10]For example, diagnosis and procedure codes would allow us to construct a measure of "severe maternal morbidity" which is defined by the Centers for Disease Control (CDC) as a set of 21 indicators for maternal morbidity https://www.cms.gov/About-CMS/Agency-Information/OMH/equity-initiatives/rural-health/rural-maternal-health. There is overlap between the CDC definition of severe maternal morbidity and our (revised) measure of maternal morbidity constructed from the NVSS data: both measures include blood transfusions and hysterectomy. The CDC measure additionally includes many extreme outcomes such as acute myocardial infarction (heart attack) and acute renal failure, whereas our NVSS-based measure includes less extreme (but more birth-related) outcomes such as 3rd/4th degree perineal lacerations.

[11]To our knowledge, such data at the national level do not exist. However, data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID) could allow for such an analysis for a limited number of states and years.

(0/1) equal to one if there was a NICU in the mother's county of birth occurrence. 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). Our Hospital Compare quality metric is an index (z-score) representing the average quality of all hospitals within a mother's county of birth occurrence. 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 1.

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:

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

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. We analyze 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.[12] 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 (National Center for Health Statistics, 2013).[13] 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. Finally, the standard errors are clustered at the county-level.

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

[12]Time-varying county-level covariates come from the National Institute for Health Surveillance, Epidemiology and End Results (SEER) Program (Surveillance, Epidemiology, and End Results, 1989-2019) and from the Regional Economic Information System (REIS) (Regional Economic Information System, 1969-2019).

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

counties have smaller and less urban populations (Table 1). 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 (without subsequent reopening) 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 (see Section A.3.1 for more detail on how we implement the propensity weighting). We find much more limited evidence of imbalance when using these weights.[14] Our main results are similar across weighted and unweighted specifications (Tables A3 to A5), suggesting any imbalance, if it exists, has minimal effects on our estimates. We also find similar results using more parsimonious versions (e.g., excluding time-varying covariates, using year fixed effects in place of δ_{uy}) and richer versions (e.g., including state-by-year fixed effects) of Eq. (1) (Tables A3 to A5).

Third, we present our main results in an event study framework; the details of the specification are discussed in Section A.3.2. 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 timing of treatment. The nature of the treatment is such that we expect the impacts to materialize immediately if the estimated relationship is causal.[15]

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

[14]In addition to the fertility rate and maternal characteristics, we also investigate whether trends in economic conditions could be influencing our results by estimating similar balance tests (and event-studies) for the employment-population ratio, per-capita earnings, and per-capita government transfers (Figure A3). This analysis reveals some evidence of increasing government transfers in treatment relative to control counties in the unweighted specification, however evidence of this trend disappears in the propensity-weighted specification.

[15]To explicitly focus on the immediate impacts of closure in estimating average (rather than dynamic) effects, we estimate an alternative specification in which the sample for treated counties is limited to a five-year window around the year of closure (i.e., two pre-closure years, the year of closure, and two post-closure years). Table A6 shows that these estimates are similar for all outcomes, suggesting that differential long-term trends are not a substantial source of bias in our main estimates.

Borusyak et al. (2021) imputation-based event study estimates are provided in Figure A2. We discuss the issue in more detail in Section A.3.3 and show that this type of bias is minimal in our setting.

Our main analysis sample excludes 1,383 counties for three reasons. First, 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 32 counties that experience only an opening. Second, we exclude 886 counties that never had an operational OB unit as the inclusion of these "always-treated" counties can exacerbate negative weighting concerns in TWFE specifications. Third, to ensure that our estimates are not picking up spillover effects, we drop 348 "receiving counties" in which at least 30% of mothers from any closure county gave birth in the three years following closure. While these restrictions are theoretically important, we find similar results in alternative models that include all of these counties and allow the treatment status to change more than once (Table A7).[16]

5 Results

5.1 Main Results

In Figures 2–5, we present event studies and the corresponding average treatment 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.

Figure 2 begins by analyzing location of birth and prenatal care. 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 27.5 percentage points (pp) on a base of 28.3%.[17] Figure 2B reveals a decline in the share of births occurring in a hospital, though the magnitude is minuscule and it is not statistically different from zero at conventional levels (-0.14pp on a base of 98.7%, p-value=0.108). Together, Figure 2A and B indicate that after a closure occurs, nearly all births

[16] We also exclude Alaska and Hawaii due to their unique geographical characteristics, and D.C. and Virginia due to their unique county definitions and frequently changing county borders in Virginia.

[17] Figure 2A shows a downward trend in the outcome in the years prior to closure, however the magnitude of the trend is small in comparison to the abrupt shift at the time of treatment. To abstract from these long-term pre-treatment trends, we use an approach similar to Finkelstein et al. (2016) and estimate a specification that limits the sample to years immediately surrounding the closures (Table A6). This alternative specification yields a treatment effect estimate of slightly smaller magnitude (22.2pp decline).

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 3 presents results for a range of infant health outcomes, and in particular Figure 3A-E present results for "upstream" outcomes that are primarily determined by conditions prior to the onset of labor (e.g., gestational age and birthweight). Figure 3A shows that closures lead to a statistically significant decline in gestational age (-0.047 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 3B shows no impact of closures on premature births, while Figure 3C shows that the gestation effect is driven by an increase in births between 37 and 39 weeks (1.4pp, p-value<0.001). Given that we find a significant decline in gestational age, we expect to see some decline in birthweight.[18] Figure 3D shows a statistically insignificant decline in birthweight in the TWFE specification (-0.86 grams), but a significant decline in the dCDH specification (-8.62 grams) that is broadly in line with a mechanical reduction in birthweight resulting from shorter gestational age.

What drives the shifting of births to 37-39 weeks? One possibility is that this stems from fewer prenatal visits (and worse prenatal health), however an alternative possibility is that births are increasingly scheduled in response to closures to avoid travel during natural labor. Figure 4A reveals that induced births at 37-39 weeks increase (1.7pp, p-value<0.001) by a similar magnitude as total births at 37-39 weeks (1.4pp, from Figure 3C). Hence, the entire effect on births at 37-39 weeks can be explained by increased inductions. Table 2 breaks these effects down by single weeks of gestational age (37, 38, and 39 weeks), showing that for both total births and inductions, births at 37, 38, and 39 weeks account for roughly 9%, 26%, and 65% of the total increase at 37-39 weeks, respectively. Taken together, this evidence suggests that providers schedule inductions to avoid long travel at the time of naturally occurring labor, especially at 39 weeks of gestation. This is consistent with recommendations from the American College of Obstetricians and Gynecologists, which lists living far from the hospital as a reason to consider elective induction at 39 weeks.[19]

The shifting of births via induction to 37-39 weeks has potential health implications. The majority of these new inductions occur at exactly 39 weeks—a point at which induction is not likely

[18] A regression of birthweight on weeks gestation in the microdata reveals that each week of gestational age is associated with 126.6 grams additional weight. Extrapolating this, a decline in 0.047 weeks would be expected to reduce birthweight by approximately 5.89 grams.

[19] See: https://www.acog.org/womens-health/faqs/labor-induction.

to be harmful, and in fact induction at 39 weeks may improve infant health relative to expectant management for low-risk mothers (Grobman et al., 2018). The health impacts of shifting births to 37 or 38 weeks (i.e., "early-term" births) are less clear. Early-term births in general are correlated with adverse neonatal outcomes such as increased use of mechanical ventilation and hypoglycemia (Sengupta et al., 2013), and even long-run outcomes such as increased obesity (Levy et al., 2017). However, these correlations may not speak to a setting such as ours where births are intentionally shifted via induction. Indeed, elective early-term induction has been shown to be uncorrelated with adverse neonatal outcomes (Salemi et al., 2016), and recent evidence finds that births that are shifted earlier by up to two weeks due to holidays are not associated with a range of negative infant health outcomes (Jacobson et al., 2021). Given this evidence, it is unlikely that the small increase in early-term births would have meaningful impacts on maternal and infant health outcomes. This is further supported below as we find no adverse impacts of closures on a range of health outcomes.[20] That said, we cannot directly rule out the possibility that the observed increase in early-term inductions leads to adverse impacts on outcomes which we cannot measure (e.g., long-term obesity).

Figure 4B presents the effects of closures on C-sections, another birth procedure that could plausibly be scheduled in response to long travel distances. We find no evidence that C-sections increase in response to closures. In fact, we find that closures lead to a clear and substantial decline in C-sections (-1.1pp, p-value<0.001). In the following section we unpack the mechanisms underlying this result and attempt to draw welfare implications.

We next turn to a range of "downstream" infant and maternal health outcomes which are plausibly 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 time frame for 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. Figure 3F-H present results for three downstream infant outcomes, and Figure 5 present results for two downstream maternal outcomes.

Figure 3F presents estimates for the most severe outcome: infant mortality. The TWFE estimate yields a null effect on infant mortality, and the 95% confidence interval excludes positive effects exceeding 5.4%. 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 at the 5% level. However, this appears

[20] Sengupta et al. (2013) find that mechanical ventilation, one of the outcomes that we observe, more strongly correlates with early-term birth than any other outcome that they measure. We find that closures lead to no detectable increase in mechanical ventilation (Figure A4).

to be anomalous. The de Chaisemartin and D'Haultfoeuille (2020) average effect estimator 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 an estimate very close to zero. Various alternative specifications also yield no effect (Table A4 and Figure A2).[21]

Figure 3G and H present results for two less severe measures of infant health. Figure 3G reveals that the share of infants with low 5-minute Apgar scores—a standard infant health measure available for the entire sample period—experiences a slight downward shift coinciding with the timing of closures, suggesting an improvement in infant health. While the magnitude of this estimate is consistent across specifications and it is statistically significant at the 5% level in the main specification (p-value=0.041), its precision varies across specifications and thus we hesitate to interpret this as meaningful evidence of an improvement in infant health. Figure 3H analyzes a composite measure of infant morbidity composed of variables available in state-years using the unrevised birth certificates, and reveals no significant impacts.

Figure 5 presents estimates for two composite measures of maternal morbidity.[22] Figure 5A uses a set of variables available in state-years using unrevised birth certificates while Figure 5B uses a more comprehensive set of maternal morbidity measures that were introduced in 2009 for states using revised birth certificates. While we observe no statistically significant impact on the unrevised measure, there is a robust decrease (improvement) in the revised maternal morbidity measure coinciding precisely with the timing of treatment (p-value = 0.001). The magnitude implies an improvement in maternal morbidity by approximately 1.5% of a standard deviation, on average.[23] The improvement in maternal morbidity can largely be attributed to reductions in maternal blood transfusions and perineal lacerations. Figure A4 provides estimates for all components of the composite measures, and reveals that maternal transfusions decline by 0.19pp on a base of 0.7%, and perineal lacerations decline by 0.28pp on a base of 1.1%. Notably, perineal lacerations are heavily concentrated among vaginal births, meaning the overall improvement in maternal morbidity is unlikely attributable to the observed decline in C-section rates (since lower C-section rates would be predictive of an increase in perineal lacerations). This is supported further by more di-

[21] In addition to varying the covariates and fixed effects, Table A4 also presents results for neonatal mortality (death within 28 days of birth). The estimate for neonatal mortality is negative (coefficient = -0.076) and more precisely estimated (standard error = 0.142).

[22] We also examine maternal mortality and present these estimates in Figure A5. Maternal deaths are fortunately a rare outcome (one maternal death occurs for every 53 infant deaths), and especially so in the set of relatively low-population counties that experience closures (40% of closure counties never experienced a maternal death in our 31 year sample). Due to the rare nature of the outcome, our estimates too imprecise to draw any meaningful conclusions.

[23] Composite measures are constructed on the micro-level data. As such, standard deviations represent variation across individuals rather than counties (individual-level standard deviations are larger than county-level standard deviations).

rect evidence showing that the improvement in maternal morbidity is concentrated among vaginal births (Table A8). Overall, we conclude that the evidence on welfare-relevant measures of infant health (Figure 3) and maternal health (Figure 5) suggest that the average effects of closures 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 2.1pp (Table A3). Using the estimate from Grobman et al. (2018), this implies a reduction in C-sections of approximately 0.3%, or 0.1pp. This explains about one-tenth of the overall decline in C-section delivery.

A second possible mechanism is that women are reallocated to hospitals with different C-section practices. To explore this possibility, Figure 6A tests whether closures induce women to 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).[24] 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 6A shows that closures prompt women to give birth in counties that have, on average, 0.9pp lower risk-adjusted C-section rates. In Figure 6B 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

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

share.[25]

Figure 6B 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 6A, 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.[26] Figure 6B reports that the effect of a closure on C-sections is particularly large (-2.0pp, 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). We caution against causal interpretations of these heterogeneous effects due to lack of exogenous variation in C-section gaps, however this evidence supports the hypothesis that local C-section rates are 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 decrease in C-sections as welfare-improving.[27] However, Currie and MacLeod (2017) show that health outcomes improve when C-section rates are either: decreased among mothers with a low predicted need of C-section, or increased among mothers with high need. Thus, it would inappropriate to conclude that the observed decrease in C-sections is welfare-improving if it were concentrated among high-need 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. We then estimate the effects of closures on C-sections across terciles of the need distribution. We find statistically significant declines in C-sections among all three quartiles (Table A8). A decline in C-sections among high-need mothers raises a concern that health outcomes may be negatively affected for this group, however we find no evidence of worsened infant or maternal health outcomes among high-need mothers (Table A8). It is possible that there are negative health effects present due to insufficient C-sections for high-need mothers, but that they are offset by improved health through other mechanisms such as increased quality of care. Consequently, we refrain from making welfare conclusions regarding the reduction in C-sections.

[25] We define receiving counties using *pre*-closure market share to ensure that market shares are not endogenous to treatment. However, it is possible that pre-closure market shares are not predictive of post-closure market shares if there is sufficient selection in out-of-county births prior to closure. In an alternative specification, we define receiving counties using market shares in the three years after post-closure and find very similar results, see Figure A6.

[26] 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 4B).

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

Next we focus on uncovering the mechanism for 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 likely 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. First, 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.

To further probe whether travel during labor is an important mechanism, we test whether deleterious impacts appear for closures that result in especially long travel distance. Figure 7 provides estimates for 16 outcomes from a version of Eq. (1) that replaces the closure indicator with a quadratic in distance to the nearest OB unit.[28] In many ways, this is a more general specification as it utilizes variation in distance resulting from both closures and openings (though there are few openings), it utilizes variation resulting from closures/openings in nearby counties (which may affect counties that do not have their own OB unit), and it allows for a nonlinear relationship between distance and the outcomes. Across all outcomes, we find no evidence that harmful impacts emerge at the longer distances observed in our data (i.e., the slopes do not tend to change in the health-worsening direction at long distances). The closures we observe generally do not result in extreme distances (e.g., the 95th percentile distance in the year following closure is 67.9 miles), which means our results cannot speak to the impacts of extreme travel distances on infant and maternal health. These estimates also provide a useful validation of our main results: across all outcomes for which we see a significant impact in the main specification, we observe an effect of similar magnitude and significance in this nonlinear distance-based approach as well.

In Figure 8A and Figure 8B, 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 (Agency for Healthcare Research Quality, 2014). Hospital Compare provides several quality measures, and ours is a composite of four com-

[28] Distance is measured as the straight-line distance between the population weighted centroid of the mother's county of residence and the population-weighted centroid of the nearest county with an operational OB unit.

monly used measures (Doyle et al., 2019).[29] Second, we measure OB-specific hospital resources using the presence of a NICU.[30] Both Figure 8A and Figure 8B provide clear evidence that closures, on average, prompt women to give birth in counties with higher quality hospitals. Figure 8A shows closures lead women to give birth in counties that have 0.1 standard deviations higher quality scores, and Figure 8B reveals that women are 4pp more likely to give birth in a county with a NICU.

In Figure 8C and Figure 8D, we replicate the exercise of plotting pre-closure gaps (as in Figure 6B), 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 robust and significant improvement on average, the revised maternal morbidity composite (i.e., the outcome in Figure 5B), and test whether the effects of closures are heterogeneous across the distribution of quality gaps. Figure 8C 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 capture 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 8D 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 (1.54% of a 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 about 70% larger compared to the effect for closures with a below-median NICU gap (0.89% of a standard deviation decrease), the difference is not statistically significant. This emphasizes the difficulty in precisely measuring the returns to quality in health care given that 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 further 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-

[29] 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 A4.

[30] 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.

infant death data to construct a predicted probability of infant death for every birth. We then estimate a "first-stage" regression across risk groups (the outcome is the share of births in the mother's county of residence, as in Figure 2A). Figure 9 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 less than 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. Indeed, in the year prior to closure, 84.3% of mothers in the highest risk groups were already traveling outside their county to give birth. Consequently, as the complier mothers are less likely to have complicated deliveries, we should not expect closures to have large effects on extreme outcomes, such as infant mortality.

5.3 Heterogeneity

Given well-known disparities across demographic groups in infant and maternal health, it is natural to ask whether certain groups are differentially affected by the closures. We estimate the impacts of closures separately by race (non-Hispanic White, Hispanic, and non-Hispanic Black), education (no college, some college or more), and mother's age (under 25, 25-34, 35+). In Table 3, we present heterogeneous effects of closures on location of birth, prenatal care, and characteristics of the birth environment. The main takeaway of this analysis is that the impacts of closure tend to be larger for more disadvantaged groups. For example, estimates for the "first-stage" outcome (birth in the mother's county of residence) reveal that Black/Hispanic, low-education, and young (<25) mothers are much more likely to be treatment compliers relative to their counterparts. The estimates for follow-on outcomes tend to mirror the patterns observed in the first-stage. The remainder of Table 3 reveals that the same groups experience larger decreases in prenatal visits, are more likely to be reallocated to counties with lower C-section rates, and are more likely to be reallocated to counties with higher quality hospitals and NICUs.

We also present heterogeneous effects for a range of infant and maternal health outcomes in Tables A9 to A11. For these health outcomes, we lack statistical power to make precise comparisons across groups, however there are at least two points worth noting. First, the observed improvement in maternal morbidity appears to be concentrated among the disadvantaged groups (although we cannot reject equal impacts at the 5% level). Second, we continue to find little evidence of deleterious health impacts across all subgroups.[31]

5.4 Empirical Concerns and Robustness

We observe that closures lead to a small increase in out-of-hospital births. If out-of-hospital births are systematically correlated with birth complications (e.g., maternal morbidity) or measures of health at birth (e.g., Apgar scores), then our estimates for these outcomes could be biased. Recall that our estimate for the effect of closures on out-of-hospital births is small in magnitude—0.14 percentage points (Figure 2)—and statistically insignificant, which indicates that the scope for misreporting bias to impact our estimates is limited. Nevertheless, we investigate how large this reporting bias would have to be to impact our estimates. First, consider the revised measure of maternal morbidity—an outcome for which we find a statistically significant improvement. The composite measure is standardized and thus makes magnitude comparisons difficult, but Figure A4 displays results for each component of the composite measures. Figure A4 shows that closures lead to significant declines in maternal transfusions (0.19pp on a mean of 0.7%) and perineal lacerations (0.28pp on a mean of 1.1%); each of these effects is larger than the change in out-of-hospital births. As such, for misreporting to explain the effects on these outcomes, we would have to assume that essentially all out-of-hospital births result in these complications, which is implausible given the mean of each outcome. Next consider measures of infant health. Our main specification reports a (marginally significant) decline in low Apgar scores (0.26pp on a mean of 4.3%); again this magnitude suggests that essentially all out-of-hospital births would have to result in low Apgar scores to explain this effect. As a final illustration, consider gestational age: we find closures lead a change in average gestational age by -0.047 weeks. If all of the new out-of-hospital births over-reported gestational age by 10 weeks (i.e., implausibly large over-reporting), this would change our estimates for gestational age from -0.047 weeks to -0.034 weeks. In conclusion, misreporting from out-of-hospital births would need to be implausibly large to meaningfully affect our estimates, and thus, we believe that possible data misreporting due a slight rise in out-of-hospital births has little impact on our final conclusions.

[31] There are two estimates that are positive and significant at the 5% level: prematurity for the age <25 group (p-value=0.047) and infant morbidity for the some college or more group (p-value=0.049). However, in this heterogeneity analysis we are estimating impacts for 8 groups and 17 outcomes (136 coefficients) and as such we expect a number of type I errors to occur by chance. Since these are both marginally significant with p-values very close to 0.05, any formal correction for multiple hypothesis testing would render these insignificant. The same cannot be said for evidence of health improvements: for the revised maternal morbidity measure, the p-values for the age <25 group and no college education group are 0.000063 and 0.000060, respectively. These survive even the most conservative correction for multiple hypothesis testing: correcting a 0.05 error rate for 136 hypothesis tests using the Bonferroni method yields a corrected error rate of 0.00037.

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 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.

Our work fits with a broader literature within health economics examining provider exits in health care, including physicians (Sabety, 2022), nursing homes (Olenski, 2022), family planning clinics (Fischer et al., 2018), and hospitals (Carroll, 2022; Gujral and Basu, 2019; Avdic, 2016; Buchmueller et al., 2006). Most related to our own work, the hospital closure literature highlights two important sources of heterogeneity in the effects—the urbanicity of the area and the urgency of health care. Urban closures lead to a relocation of care outside of the hospital setting and negligible impacts on time sensitive conditions (e.g., acute myocardial infarction, stroke) (Gujral and Basu, 2019; Buchmueller et al., 2006). On the other hand, rural closures for time sensitive conditions are associated with elevated mortality risk (Carroll, 2022; Gujral and Basu, 2019). The distinction of our results with this collective work is not surprising given that perinatal care involves considerable advance planning and is less time sensitive when compared to acute myocardial infarction and stroke. Together, these papers in conjunction with our own highlight that closures can have very different health impacts depending on the types of outcomes and populations under study.

References

Agency for Healthcare Research Quality, 2014. Hosptial Compare Quality Ratings."https://www.ahrq.gov/data/monahrq/data/index.html".

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.

Almond, Douglas, Janet Currie, and Valentina Duque, "Childhood circumstances and adult outcomes: Act II," *Journal of Economic Literature*, 2018, *56* (4), 1360–1446.

American Hopstial Association, 1995-2016. American Hospital Association (AHA) Annual Survey Database.

Avdic, Daniel, "Improving efficiency or impairing access? Health care consolidation and quality of care: Evidence from emergency hospital closures in Sweden," *Journal of Health Economics*, 2016, *48*, 44–60.

_ , Petter Lundborg, and Johan Vikström, "Does healthcare consolidation harm patients? Evidence from maternity ward closures," *American Economic Review: Economic Policy*, forthcoming.

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.

Buchmueller, Thomas C, Mireille Jacobson, and Cheryl Wold, "How far to the hospital?: The Effect of hospital closures on access to care," *Journal of Health Economics*, 2006, *25* (4), 740–761.

Carroll, Caitlin, "Impeding access or promoting efficiency? Effects of rural hospital closure on the cost and quality of care," *Working Paper*, 2022.

_ , Julia D Interrante, Jamie R Daw, and Katy Backes Kozhimannil, "Association between medicaid expansion and closure of hospital-based obstetric services: Study examines the association between Medicaid expansion and the closure of hospital-based obstetric services.," *Health Affairs*, 2022, *41* (4), 531–539.

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.

Fischer, Stefanie, Heather Royer, and Corey White, “The impacts of reduced access to abortion and family planning services on abortions, births, and contraceptive purchases,” *Journal of Public Economics*, 2018, *167*, 43–68.

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.

Gujral, Kritee and Anirban Basu, “Impact of rural and urban hospital closures on inpatient mortality,” 2019.

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.

Jacobson, Mireille, Maria Kogelnik, and Heather Royer, "Holiday, just one day out of life: Birth timing and postnatal outcomes," *Journal of Labor Economics*, 2021, *39* (S2), S651–S702.

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.

Levy, Dorit Paz, Eyal Sheiner, Tamar Wainstock, Ruslan Sergienko, Daniella Landau, and Asnat Walfisch, "Evidence that children born at early term (37-38 6/7 weeks) are at increased risk for diabetes and obesity-related disorders," *American Journal of Obstetrics and Gynecology*, 2017, *217* (5), 588–e1.

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.

Lorch, Scott A, Sindhu K Srinivas, Corinne Ahlberg, and Dylan S Small, "The impact of obstetric unit closures on maternal and infant pregnancy outcomes," *Health Services Research*, 2013, *48* (2pt1), 455–475.

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

National Center for Health Statistics, 1989-1991 & 1996-2017. Data File Documentations, Cohort Linked Births/Infant Deaths Files, National Center for Health Statistics, Hyattsville, Maryland.

_ , 1989-2019. Data File Documentations, Natality Files, National Center for Health Statistics, Hyattsville, Maryland.

_ , 1989-2019. Data File Documentations, Detailed Mortality Files, National Center for Health Statistics, Hyattsville, Maryland.

_ , 2013. "https://www.cdc.gov/nchs/data_access/urban_rural.htm".

Olenski, Andrew, "Reallocation and the (in)efficiency of exit in the U.S. nursing home industry," 2022.

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.

Regional Economic Information System, 1969-2019. "https://www.bea.gov/data/economic-accounts/regional".

Sabety, Adrienne, "The value of relationships in health care," 2022.

Salemi, Jason L, Elizabeth B Pathak, and Hamisu M Salihu, "Infant outcomes after elective early-term delivery compared with expectant management," *Obstetrics & Gynecology*, 2016, *127* (4), 657–666.

Sengupta, Shaon, Vivien Carrion, James Shelton, Ralph J Wynn, Rita M Ryan, Kamal Singhal, and Satyan Lakshminrusimha, "Adverse neonatal outcomes associated with early-term birth," *JAMA pediatrics*, 2013, *167* (11), 1053–1059.

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

Surveillance, Epidemiology, and End Results, 1989-2019. SEER*Stat Database: Incidence - SEER Research Data, 8 Registries, Nov 2021 Sub (1975-2019) - Linked To County Attributes - Time Dependent (1989-2019) Income/Rurality, 1989-2019 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2022, based on the November 2021 submission.

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.

US Census Bureau, 2010. US Census Bureau Urban and Rural."https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html".

US Census Bureau County and State Shapefiles, 2014. "https://www2.census.gov/geo/tiger/TIGER2014/".

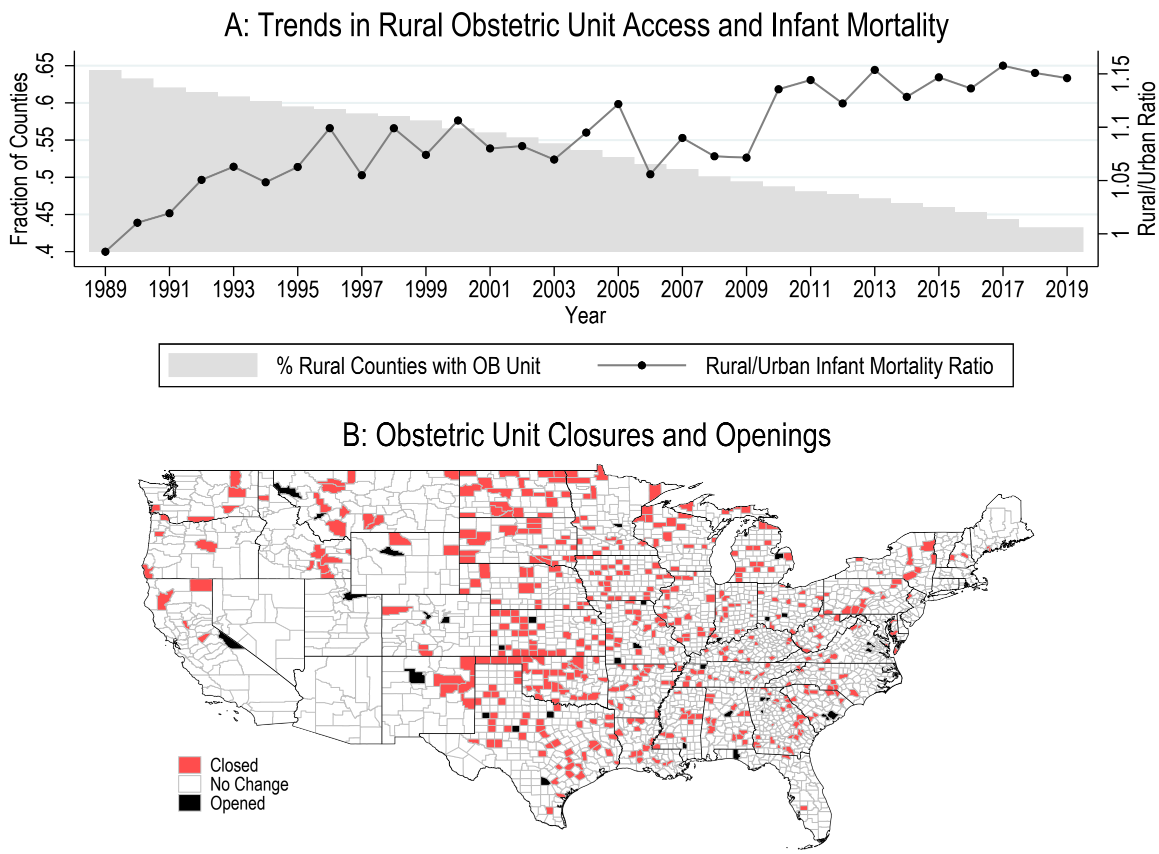
Wennberg, John and Alan Gittelsohn, "Small area variations in health care delivery: A population-based health information system can guide planning and regulatory decision-making.," *Science*, 1973, *182* (4117), 1102–1108.

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.

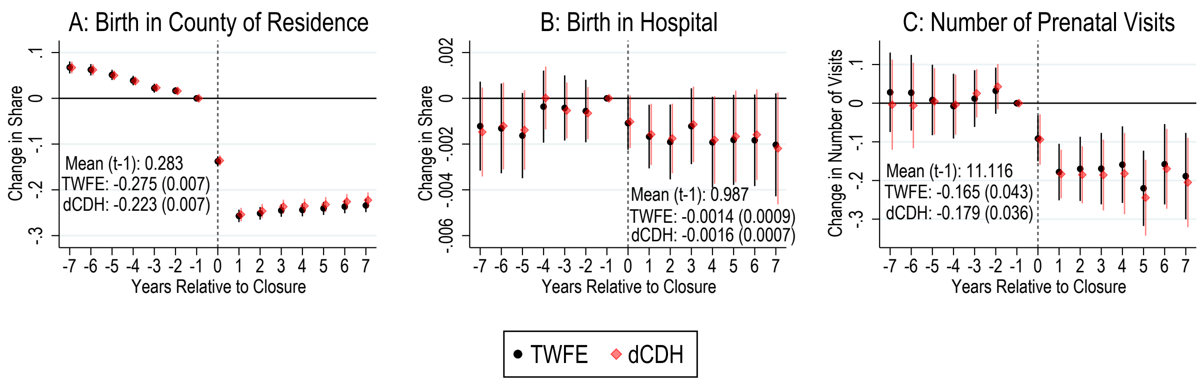
Figures & Tables

Figure 1: Temporal and Spatial Variation in OB Unit Access: 1989-2019



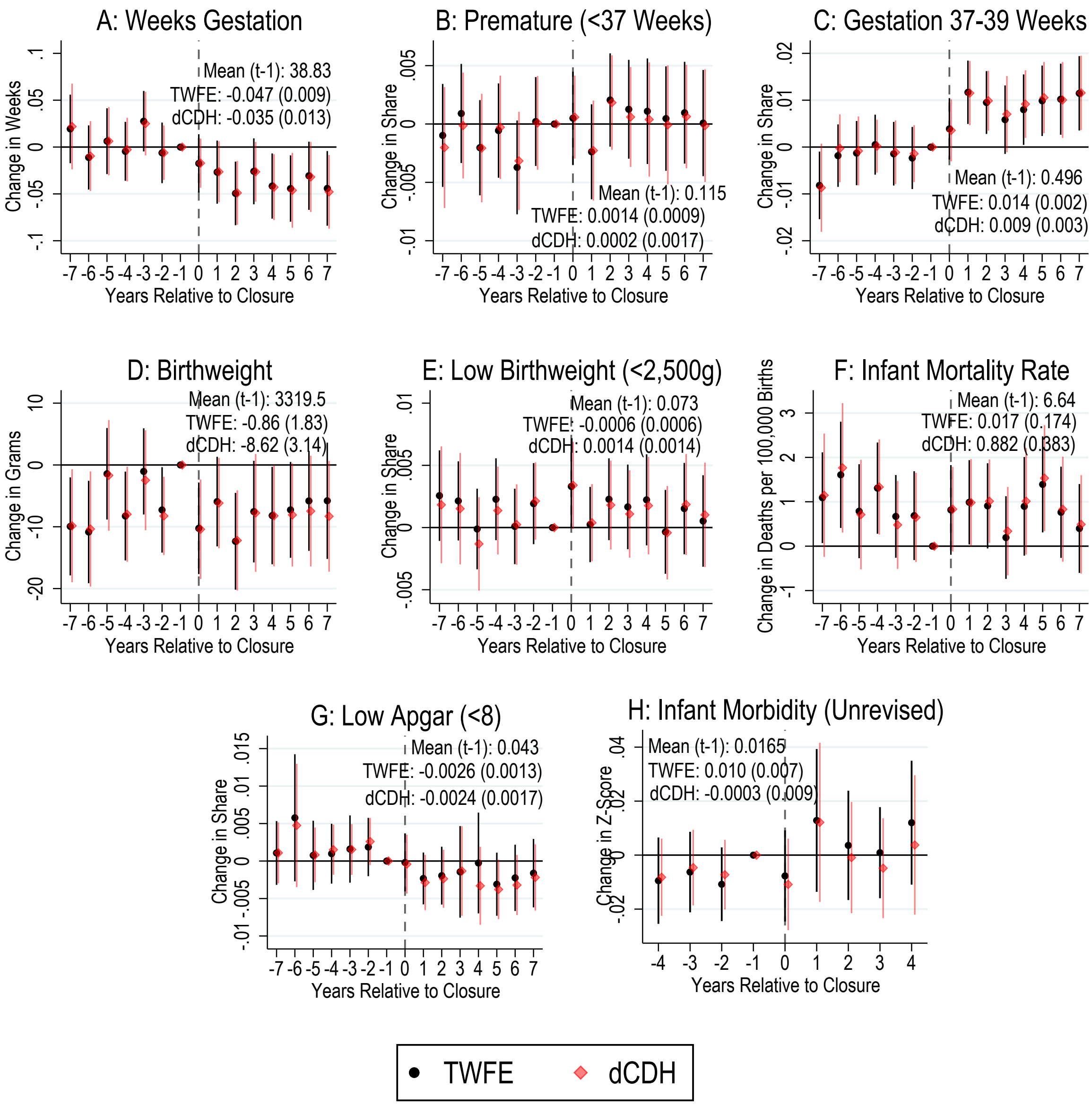
Notes: Rural counties are those classified as non-core or micropolitan in the 2013 NCHS urban/rural classification. In our sample, there are 1,883 rural counties and 1,065 urban counties. In Panel A, the shaded region displays the share of rural counties with an operational OB unit in each year. The 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 OB unit to zero, and an "opening" is the opposite. To match our main specification, counties classified as "closed" are those that experienced a closure and no subsequent reopening. Shapefiles used to construct these figures come from US Census Bureau County and State Shapefiles (2014).

Figure 2: Average Effect of Closures on Location of Birth and Prenatal Visits



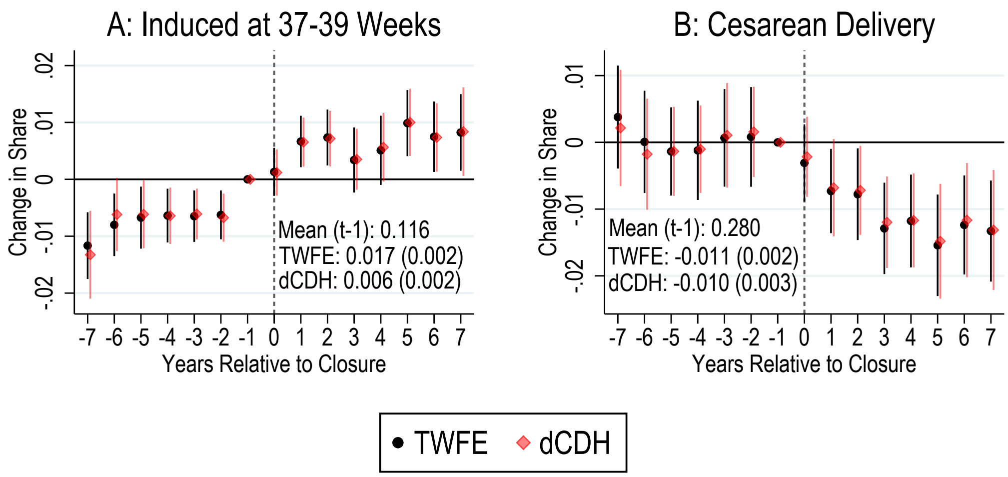
Notes: These figures plot estimated β_j from Eq. (2) using two different estimators. "TWFE" refers to estimates from a two-way fixed effects specification and "dCDH" refers to estimates from the de Chaisemartin and D'Haultfoeuille (2020) difference-in-differences estimator. Dynamic treatment effects are shown in black circles (TWFE) and red diamonds (dCDH). Bars represent 95% confidence intervals. Each subfigure also displays the mean of the dependent variable for treated counties in the year prior to closure, and the average treatment effects (standard error clustered at the county-level in parentheses) for both estimators. The average treatment effect estimate for the TWFE specification is the estimate of β from Eq. (1), and represents the average treatment effect over the entire post-treatment period. The average treatment effect for the dCDH specification essentially represents the average value of the post-treatment coefficient estimates from the event-study specification (it represents an average treatment effect for periods $t = 0$ through $t = 7$ relative to $t = -1$, rather than the average treatment effect for the entire post-treatment period relative to the entire pre-treatment period).

Figure 3: Average Effect of Closures on Infant Health



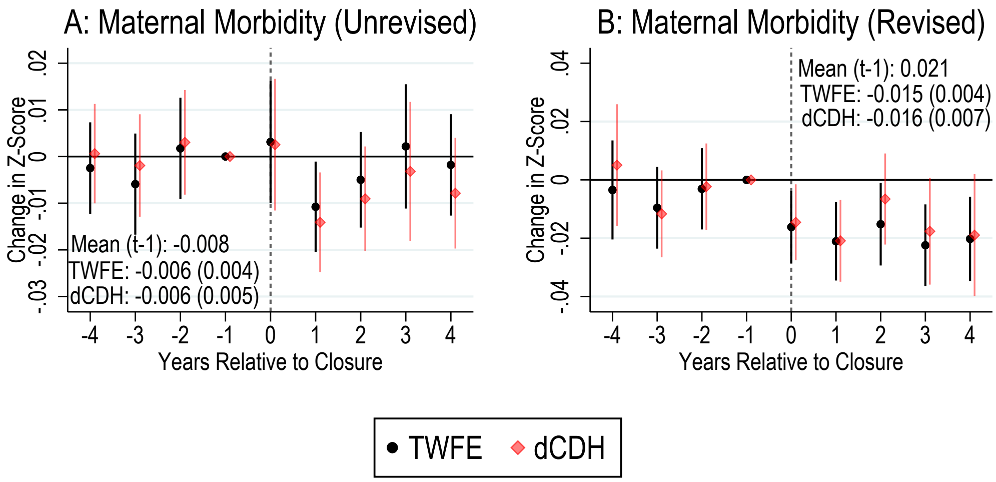
Notes: See Figure 2 for general notes on interpreting event studies. The Apgar score represents a test conducted five minutes after birth and is based on the infant's skin color, heart rate, reflexes, muscle tone, and breathing rate/effort. Scores range between 0 and 10, where 10 is the highest. The "Low Apgar (<8)" outcome represents the share of infants with a 5-minute Apgar score under 8. "Infant Morbidity (Unrevised)" is a composite outcome representing the following components: meconium staining, birth injury, infant seizures, and use of ventilator. Higher values of the composite represent worse health. Several components of these composite measures were phased out beginning with the 2003 revision of the birth certificate, and thus the measure uses only state-years using unrevised birth certificates (2006 is the most recent year all components were available in any state). Because the composite measure uses a limited sample, the event study is limited to four years pre- and post-closure.

Figure 4: Average Effect of Closures on Birth Procedures



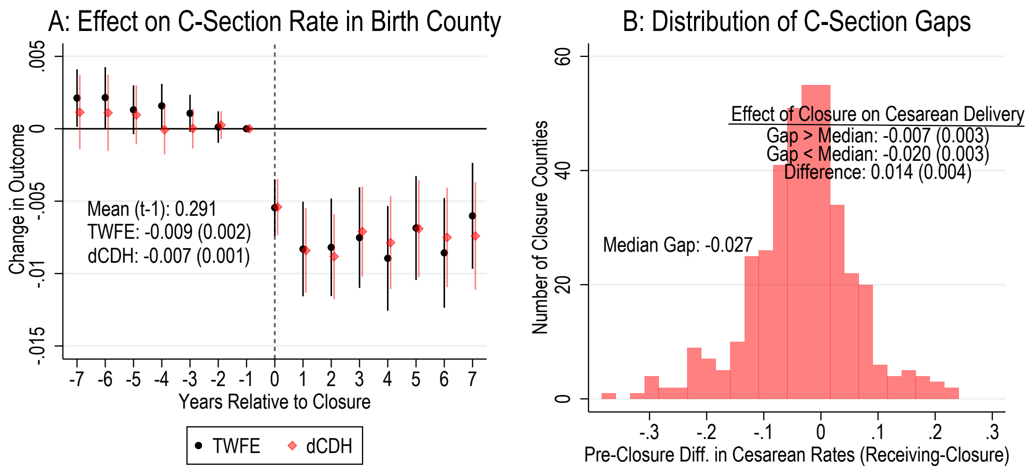
Notes: See Figure 2 for general notes on interpreting event studies.

Figure 5: Average Effect of Closures on Maternal Health



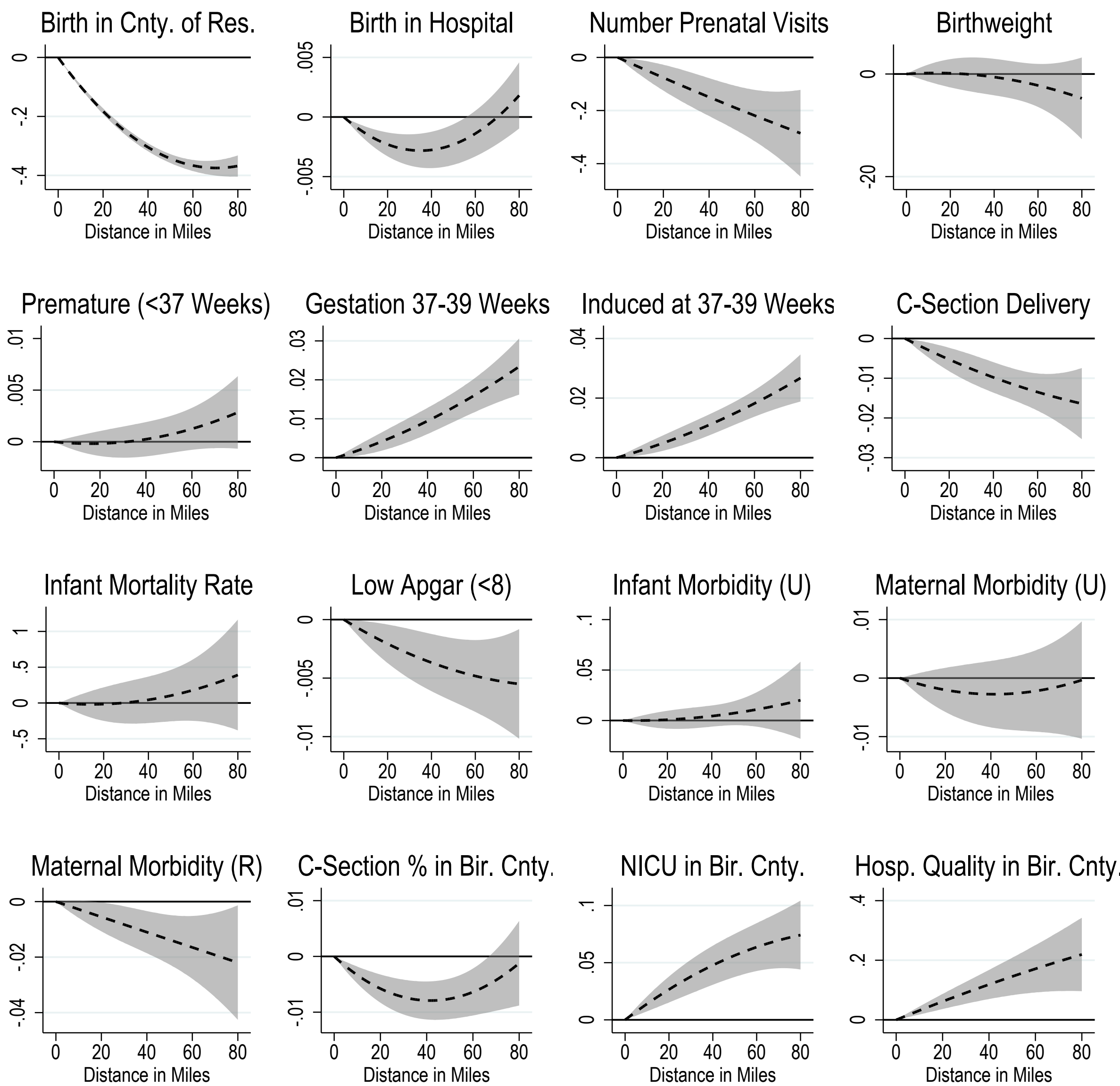
Notes: See Figure 2 for general notes on interpreting event studies. Both outcomes are composite outcomes that measure maternal morbidity, where higher values represent worse health. Several components of these composite measures were phased in or out beginning with the 2003 revision of the birth certificate, 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. "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. Because the composite measures use limited samples, the event studies are limited to four years pre- and post-closure.

Figure 6: Effect of Closures on Birth Environment (C-Section Delivery)



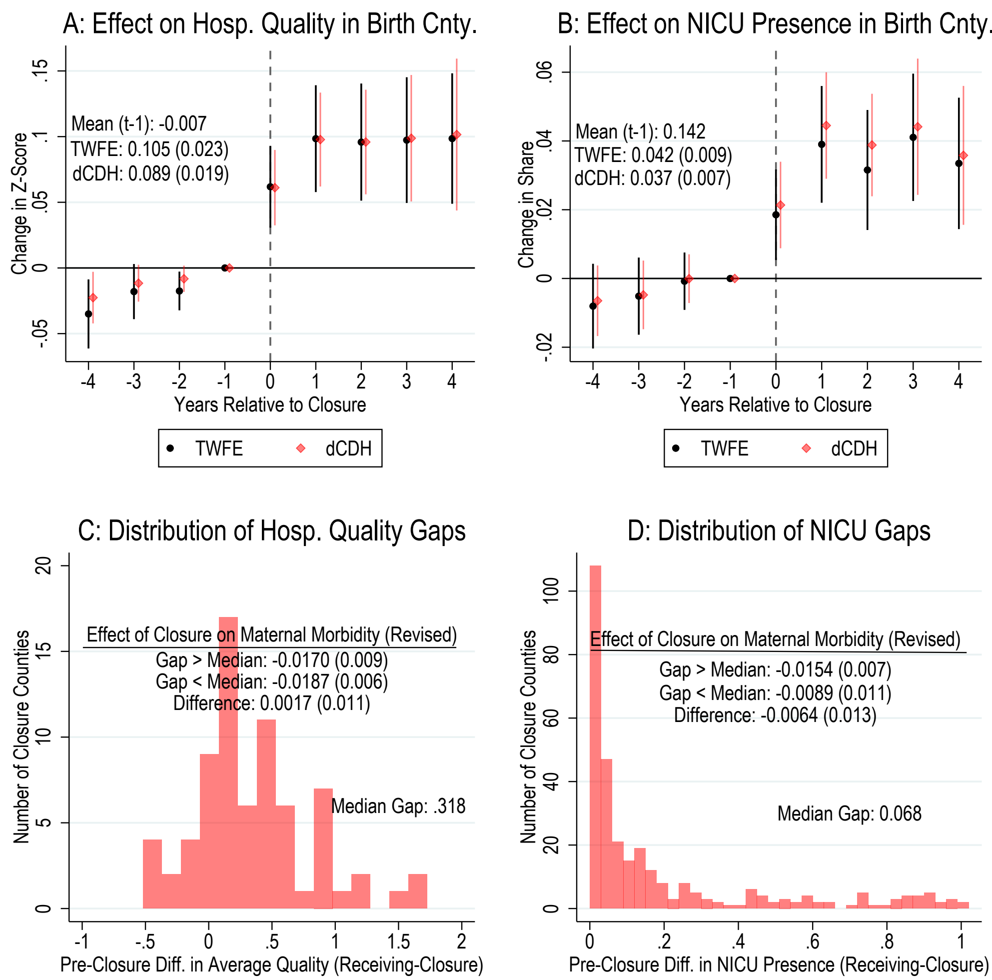
Notes: See Figure 2 for general notes on interpreting the event studies. In Panel A, the outcome for each mother is the risk-adjusted C-section rate in her county of birth occurrence in the three years prior to closure (for non-closure counties, it is a random three year period). Panel B displays the 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 labeled "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 2B; not the outcome from Figure 6A). Sample restrictions for this analysis are discussed in Section A.3.4.

Figure 7: Quadratic Distance-Based Effects of Closures



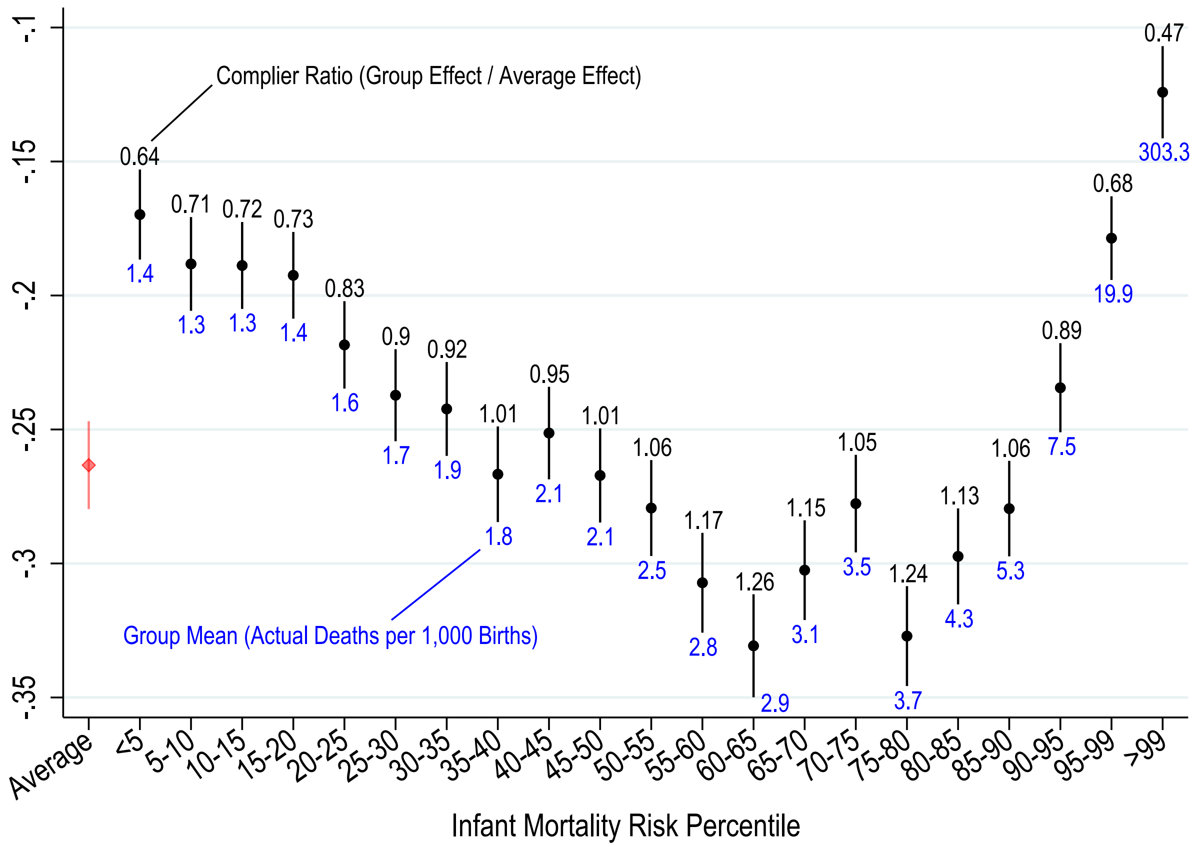
Notes: Each plot represents a separate regression in which the Closed indicator in Eq. (1) is replaced with a quadratic in the straight-line distance to the nearest OB unit. Dashed lines represent the predicted difference in the outcome between having an operational OB unit *X* miles from a mother's county of residence and having one in her county of residence (i.e., zero miles). Shaded regions represent 95% confidence intervals. The quadratic specification exploits broader variation compared to the closure indicator: 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). Accordingly, and unlike the main specification, this analysis retains counties that experience openings and counties that never had an 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 8: 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 the two hospital quality metrics are provided in Section A.2. Panels C/D display the distribution of pre-closure hospital quality/NICU gaps between the closure and receiving counties. The text labeled "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.

Figure 9: 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. This analysis requires using the linked birth-infant death files which are only available for 1989-1991 and 1996-2017 whereas other analyses of infant mortality use the unlinked mortality files available for 1989-2019.

Table 1: Mean Outcomes and County Characteristics

	All Counties	Closure Counties	Non-Closure Counties Unweighted	Non-Closure Counties P-Weighted
Panel A: County Characteristics				
Fertility Rate	66.85	67.66	66.68	67.14
Fertility Rate Growth Rate	0.008	0.0091	0.0078	0.0165
Population	94,709	22,122	109,971	22,556
Population Growth Rate	0.0049	0.0012	0.0057	0.0027
Empl./Pop.	0.511	0.491	0.515	0.489
Percent Rural	0.754	0.844	0.736	0.889
Female 15-44 Pop. Share	0.379	0.359	0.384	0.358
Panel B: Birth Location, Prenatal Care and Outcomes Determined Prior to Birth				
Occurrence in Cnty. of Res.	0.398	0.196	0.440	0.274
Occurrence in Hospital	0.985	0.985	0.985	0.984
Number of Prenatal Visits	11.20	11.04	11.24	11.08
Birthweight	3,300	3,304	3,299	3,301
Low Birthweight (<2500g)	0.076	0.075	0.076	0.075
V. Low Birthweight (<1500g)	0.0130	0.0128	0.0130	0.0125
Weeks Gestation	38.76	38.76	38.76	38.76
Premature (<37 Weeks)	0.117	0.117	0.117	0.117
Gestation 37-39 Weeks	0.510	0.511	0.509	0.514
Induced	0.239	0.245	0.238	0.246
Induced at 37-39 Weeks	0.120	0.125	0.119	0.125
Panel C: Birth Environment in County of Occurrence				
Hospital Compare Composite	0.055	0.049	0.057	0.038
NICU in Bir. Cnty.	0.412	0.384	0.418	0.336
Panel D: Health Outcomes				
Cesarean Delivery	0.278	0.282	0.277	0.280
Low Apgar (<8)	0.0421	0.045	0.042	0.044
Infant Composite (1989-2006)	0.0041	-0.0063	0.0063	-0.0128
Infant Mortality Rate	7.24	7.45	7.20	7.23
Maternal Composite (1989-2006)	-0.0012	-0.0371	0.0064	-0.0131
Maternal Composite (2009-2019)	0.0019	-0.004	0.0031	0.0146
Number of Counties	2,809	488	2,321	2,321

Counties that experience an opening at any point in the sample are excluded (as in our main specification). 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.3.1. Rural counties are those classified as non-core or micropolitan in the 2013 NCHS urban/rural classification. The largely rural nature of the closure are not immediately apparent in the summary statistics which are not population-weighted. Population weighting reveals that 12.6% of non-closure county residents reside in rural counties and 54.7% of closure county residents reside in rural counties. "U" represents measures from the unrevised birth certificates and "R" represents measures from the revised birth certificates.

Table 2: Timing of Birth and Induction

	Gestational Age			Induction at Gestational Age		
	37 Weeks	38 Weeks	39 Weeks	37 Weeks	38 Weeks	39 Weeks
Closed	0.0012	0.0035**	0.0089***	0.0015***	0.0045***	0.0111***
	(0.0007)	(0.0011)	(0.0014)	(0.0004)	(0.0007)	(0.0013)
Mean Dep. Var. (t-1)	0.084	0.161	0.251	0.018	0.035	0.062
N	48,820	48,820	48,820	48,786	48,786	48,786

Notes: Sample sizes differ because there were 34 county-years in which induction was not recorded for any birth. 29 of these observations occurred in 1989 and the remaining 5 in 1990, likely resulting from slow uptake of the 1989 revised birth certificate. Standard errors clustered at the county-level are in parentheses. ***, **, * indicate significance at the 0.1%, 1%, and 5% levels.

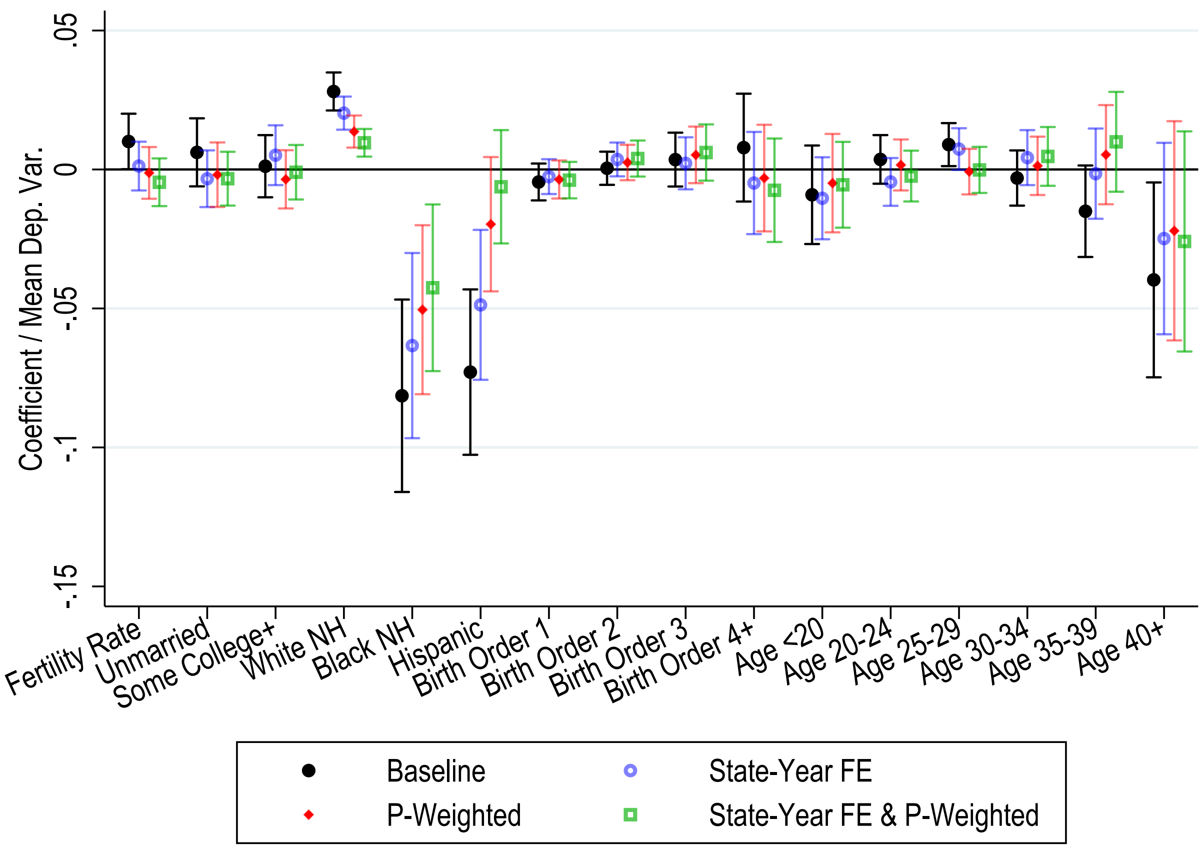
Table 3: Heterogeneous Effects of Closures on Birth Location, Prenatal Care, and Birth Environment

	Race			Education		Age		
	White NH	Hispanic	Black NH	No College	Some College+	Age<25	Age 25-34	Age 35+
Birth in Cnty. of Res.	-0.238***	-0.346***	-0.371***	-0.323***	-0.189***	-0.327***	-0.240***	-0.219***
	(0.007)	(0.011)	(0.012)	(0.008)	(0.007)	(0.008)	(0.007)	(0.007)
Difference (p-value)	-	<0.001	<0.001	-	<0.001	-	<0.001	<0.001
N	48,796	44,650	38,619	48,283	48,367	48,819	48,820	48,715
Birth in Hospital	-0.0010	-0.0016	0.0008	-0.0037*	-0.0019*	-0.0017	-0.0015	-0.0018
	(0.0010)	(0.0020)	(0.0021)	(0.0016)	(0.0008)	(0.0009)	(0.0009)	(0.0019)
Difference (p-value)	-	0.782	0.523	-	0.293	-	0.845	0.858
N	48,796	44,648	38,616	48,283	48,367	48,819	48,820	48,715
Prenatal Visits	-0.149***	-0.185*	-0.275**	-0.240***	-0.097*	-0.230***	-0.132**	-0.081
	(0.043)	(0.081)	(0.088)	(0.050)	(0.043)	(0.047)	(0.043)	(0.051)
Difference (p-value)	-	0.744	0.214	-	<0.001	-	<0.001	<0.001
N	48,796	44,555	38,500	48,264	48,352	48,815	48,820	48,710
C-Sect. % in Bir. Cnty.	-0.009***	-0.011***	-0.013***	-0.012***	-0.007***	-0.012***	-0.008***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Difference (p-value)	-	0.052	0.013	-	<0.001	-	<0.001	<0.001
N	40,369	37,456	32,622	40,332	40,337	40,321	40,337	40,188
Hosp. Qual. in Bir. Cnty.	0.017*	0.046***	0.062***	0.031***	0.012	0.032***	0.017*	0.009
	(0.007)	(0.012)	(0.015)	(0.009)	(0.006)	(0.009)	(0.007)	(0.007)
Difference (p-value)	-	<0.001	<0.001	-	<0.001	-	<0.001	<0.001
N	47,226	41,600	36,044	46,578	46,764	47,116	47,208	46,691
NICU in Bir. Cnty.	0.039***	0.057***	0.069***	0.042***	0.030**	0.050***	0.037***	0.035***
	(0.010)	(0.011)	(0.015)	(0.010)	(0.010)	(0.010)	(0.009)	(0.010)
Difference (p-value)	-	0.011	0.003	-	0.019	-	<0.001	0.003
N	34,649	32,256	27,763	34,143	34,225	34,650	34,650	34,569

Notes: Each coefficient represents a separate regression. "Difference (p-value)" represents a test of equality against the first group in the category (e.g., in the "Hispanic" column it represents a test of Hispanic against White non-Hispanic). The sample sizes vary across columns because observations with no births in a given group-county-year are dropped. Educational attainment was not measured from 2009-2013 for a small number of states that had not yet switched to revised birth certificates; as such, those observations are dropped. ***, **, * indicate significance at the 0.1%, 1%, and 5% levels.

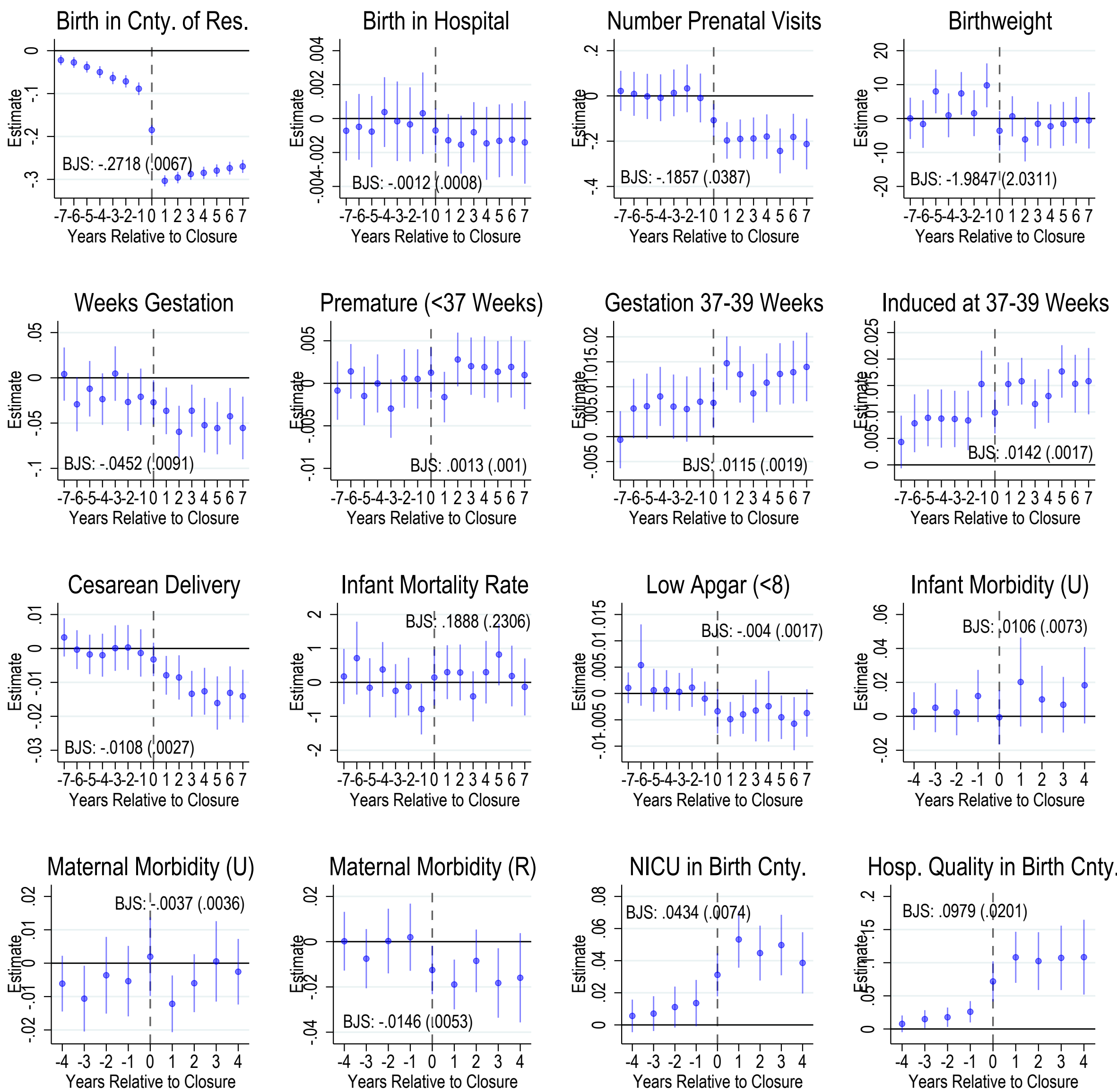
Appendix Figures

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



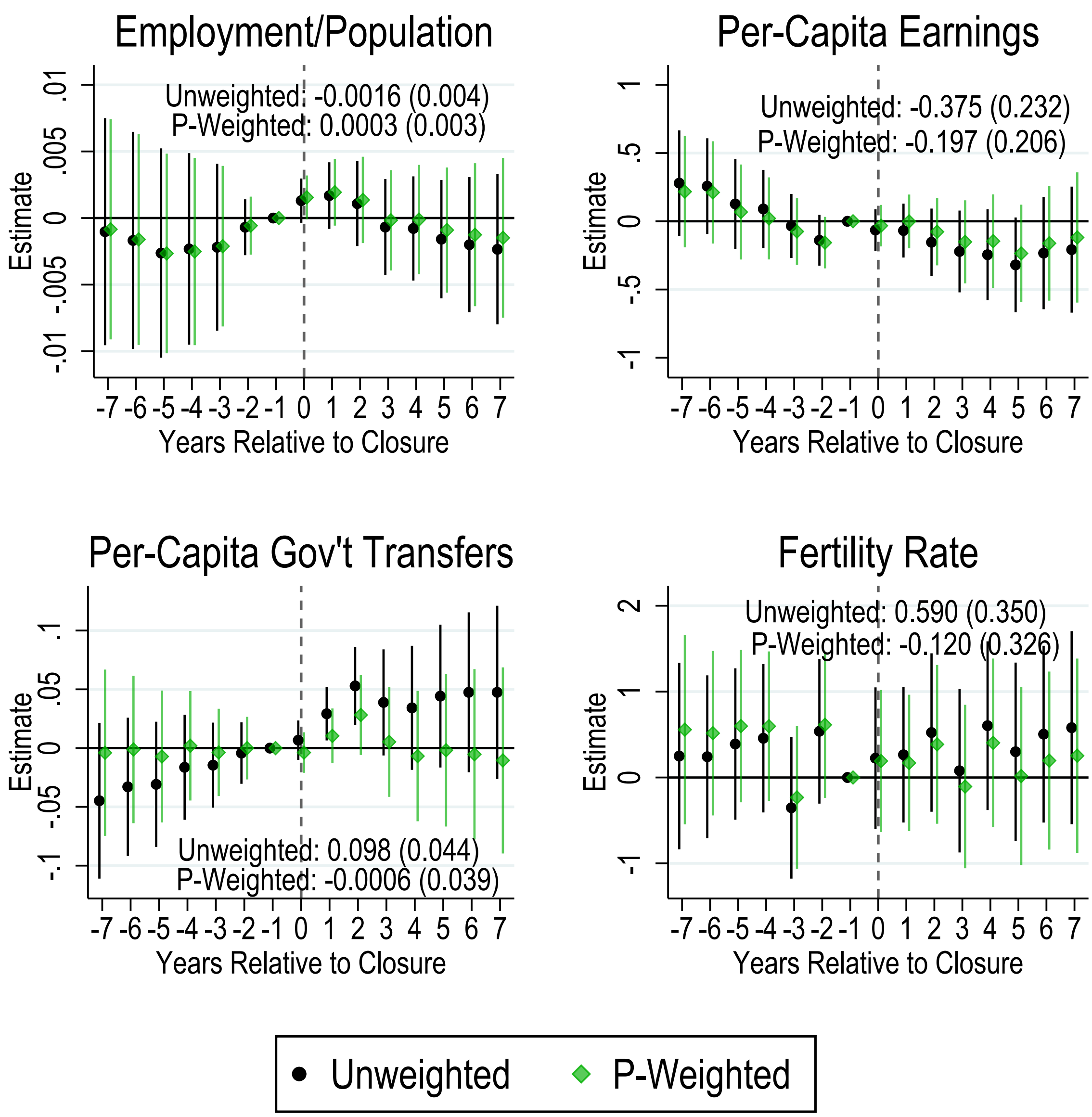
Notes: For comparability across outcomes, all coefficient estimates are divided by the mean of the dependent variable. The baseline specification (black solid circles) is described in Eq. (1). The second specification (blue open circles) adds state-by-year fixed effects. The third specification (red diamonds) weights by the propensity to experience a closure. The process of calculating propensity score weights is described in Section A.3.1. 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 open squares) includes state-by-year fixed effects and propensity weights.

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



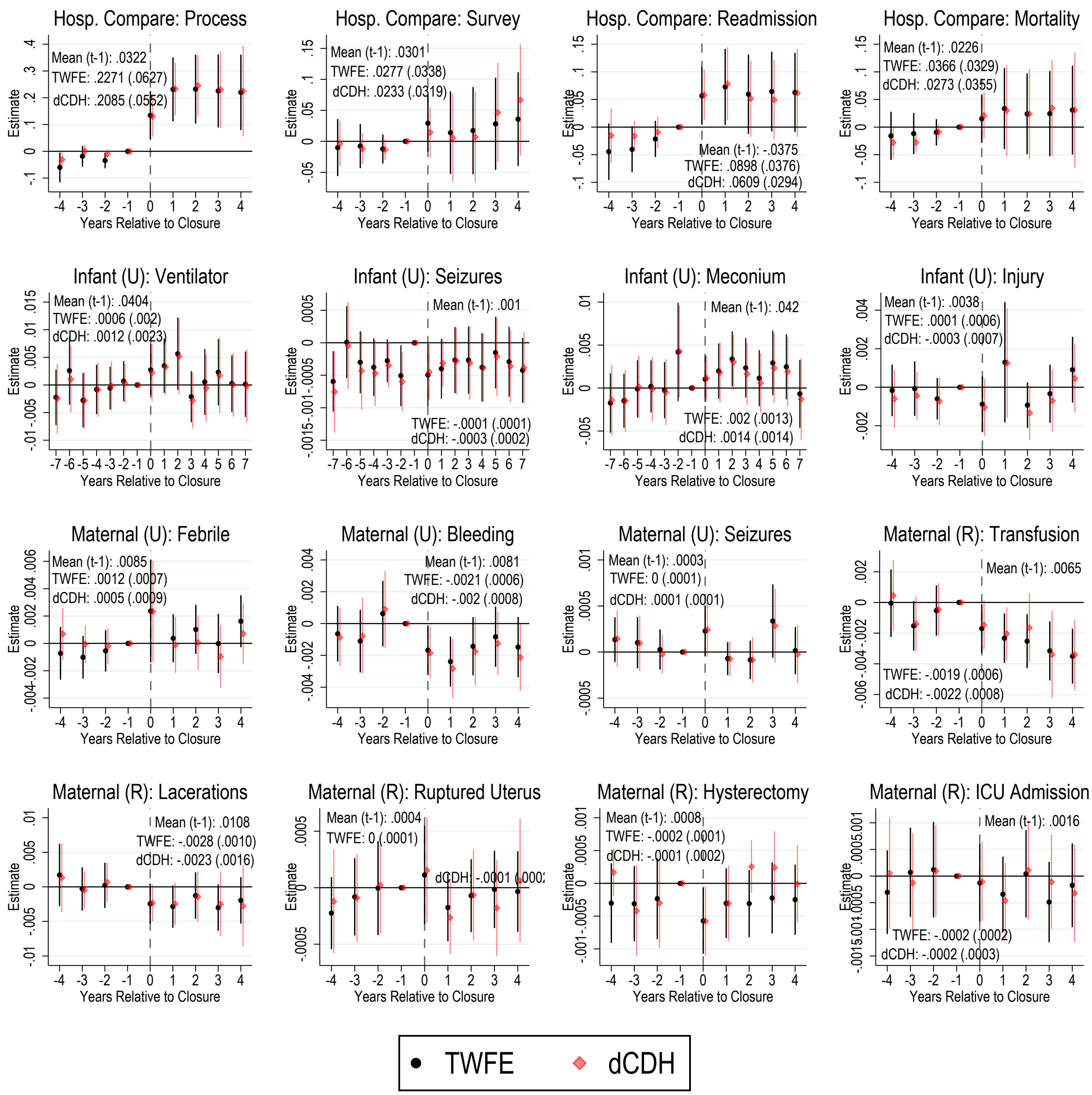
Notes: These plots replicate the 16 estimates presented in Figures 2–5 and 8 using the Borusyak et al. (2021) imputation-based difference-in-differences estimator. The point estimate labeled "BJS" on each plot represents the average effect across the post-treatment periods $t = 0$ through $t = 7$. All 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 Borusyak et al. (2021) estimator uses the following three-step imputation procedure. First, unit and time fixed effects are calculated by regressions using only untreated observations. Second, those fixed effects are used to impute untreated potential outcomes, and thereby create an estimated treatment effect for each treated observation. Third, the estimation target is calculated as an average of the treatment effect estimates. A key feature of this imputation procedure is that treatment effects for each period relative to treatment are not calculated relative to a specific pre-treatment period (typically $t-1$) as they are in a typical TWFE approach and in other newly developed DiD estimators such as de Chaisemartin and D'Haultfoeuille (2020). Instead, the imputation procedure imputes untreated potential outcomes from the full set of untreated observations and provides treatment effect estimates for every period relative to treatment including $t-1$.

Figure A3: Event Studies for Economic Variables and Fertility Rate



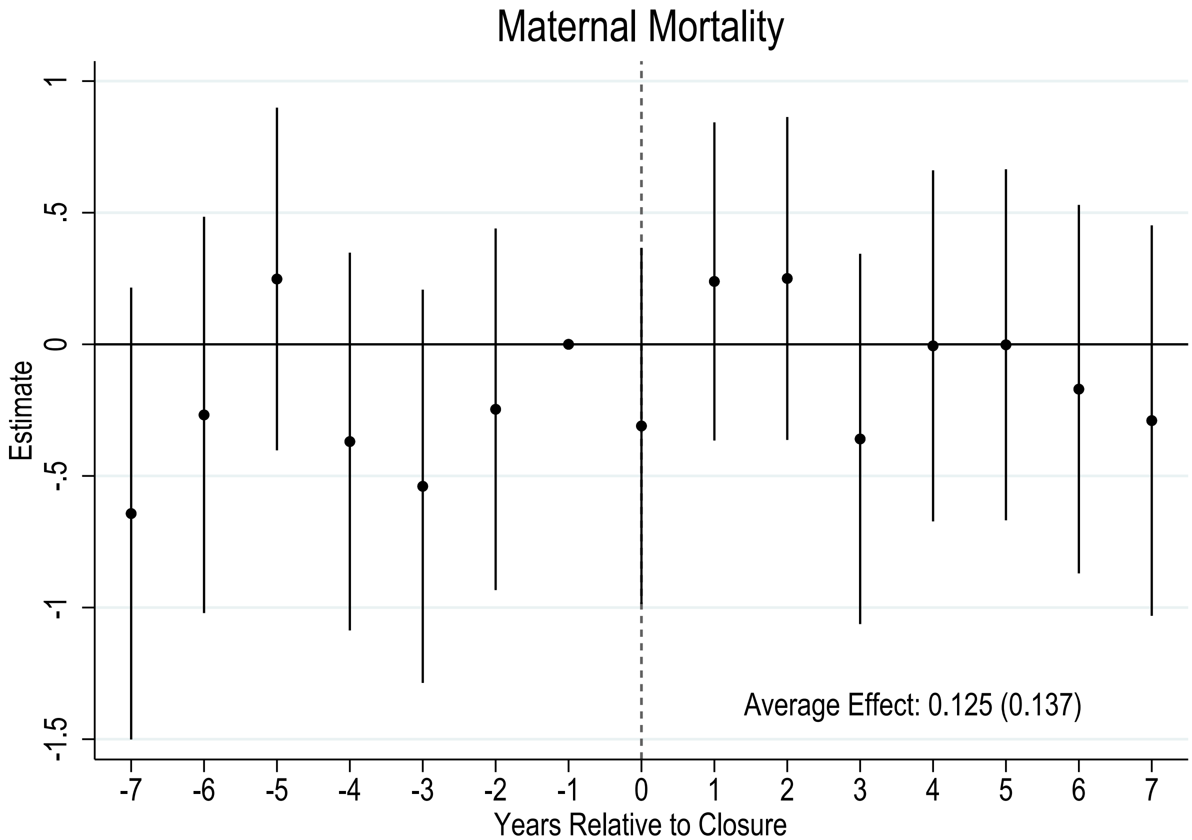
Notes: See Figure 2 for general notes on interpreting the event studies. "Unweighted" refers to our main specification, and "P-Weighted" refers to a specification in which counties are weighted by the propensity to experience a closure. The process of calculating propensity score weights is described in Section A.3.1

Figure A4: Effect of Closures on Components of Composite Measures



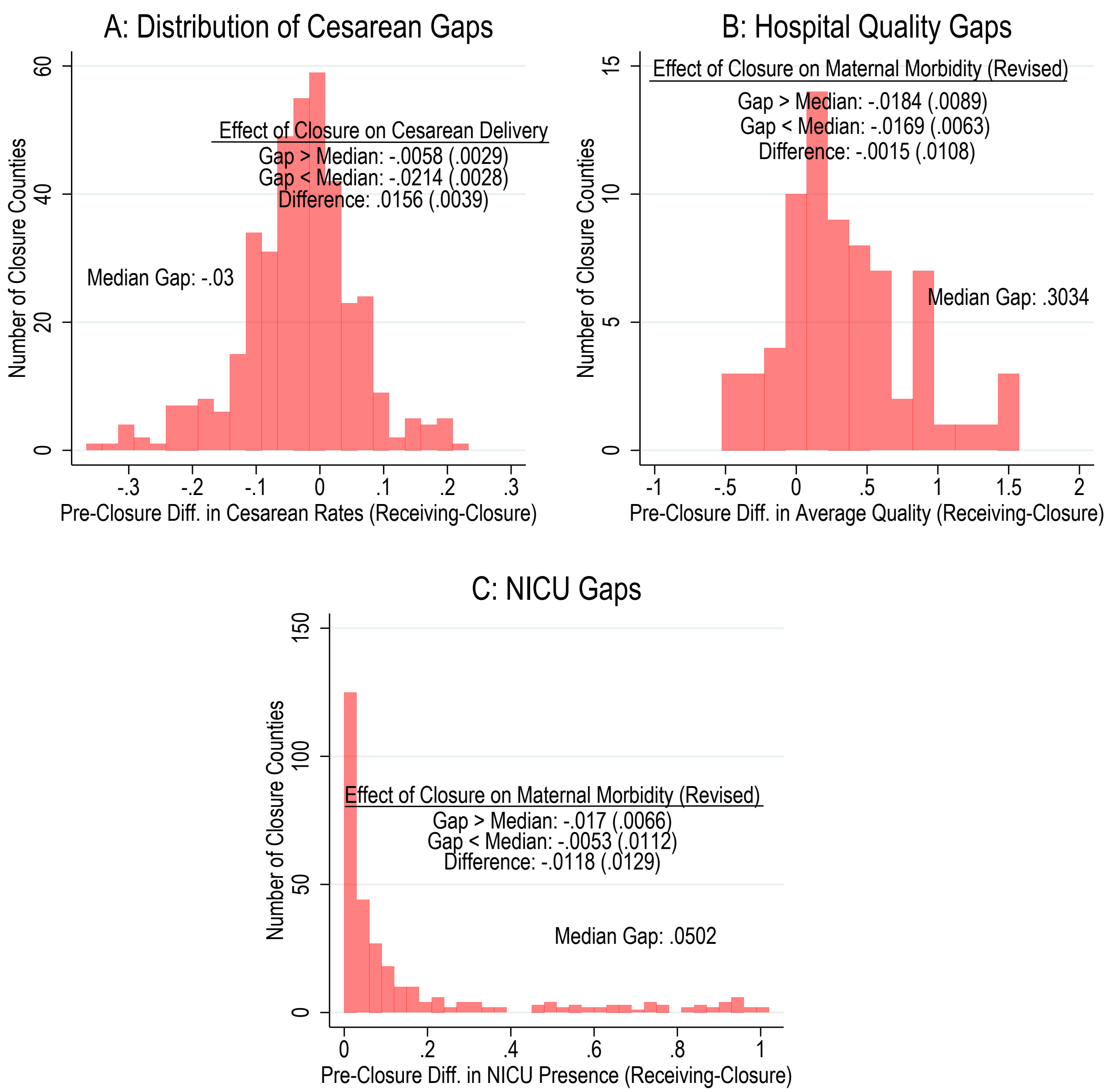
Notes: See Figure 2 for general notes on interpreting event studies. The top four plots show effects of closures on the four components of the hospital quality composite from Hospital Compare. We follow Doyle et al. (2019) in constructing the four measures: process measures, patient survey measures, 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. Table A1 details the years and the number of states for which each of these variables is available.

Figure A5: Effect of Closures on Maternal Mortality



Notes: Because maternal mortality is a rare outcome, it would be inappropriate to analyze this outcome using ordinary least squares as we do for other outcomes in this analysis. Instead, we use logistic regression and define the outcome as an indicator for any maternal deaths occurring in a given county-year. Among treated counties, 4.0% of county-year cells experienced a maternal death. 40% of treated counties *never* experienced a maternal death over our 31 year sample, and are automatically excluded from the analysis because there is no variation in the outcome within these counties. We include the same controls (fixed effects and time-varying covariates) described in Eq. (1). Estimates in the figure represent coefficients from the logistic regression. The event study reveals no visual evidence of a change in the outcome coinciding with the timing of treatment, however the estimates are extremely imprecise. The 95% confidence interval includes changes in maternal deaths ranging from -13.3% to 48.2%.

Figure A6: Alternative "Receiving" County Definition



Notes: This figure replicates Figure 6B and Figure 8C,D using an alternative definition for "receiving" counties. Specifically, here receiving counties are defined using their market share in the three years post-closure (rather than pre-closure in the main specification).

Appendix Tables

Table A1: 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	46	45	47	47	47	47	47	0	0	0	0	0
1990-1995	47	44	47	47	47	47	47	0	0	0	0	0
1996	47	44	47	47	47	46	47	0	0	0	0	0
1997-2002	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 A2: 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.	
Closed	-0.234***	-0.00181	-0.155**	-2.225	0.0000458	
	(0.00907)	(0.000932)	(0.0501)	(2.287)	(0.000894)	
N	33,968	33,968	33,968	33,968	33,968	
	Panel B: Gestation and Induction					
	Weeks Gestation	Premature (<37 Weeks)	Gestation 37-39 Weeks	Induced at 37-39 Weeks	Induced Ever	
Closed	-0.0324**	0.00132	0.0114***	0.0177***	0.0157***	
	(0.0116)	(0.00114)	(0.00246)	(0.00391)	(0.00260)	
N	33,968	33,968	33,968	33,968	33,968	
	Panel C: Maternal and Infant Health Outcomes					
	Cesarean	Low APGAR	Infant Morbid. (Unrevised)	Infant Mortality Rate	Maternal Morbid. (Unrevised)	Maternal Morbid. (Revised)
Closed	-0.0113***	-0.00223	0.00355	-0.141	-0.00406	-0.00671
	(0.00244)	(0.00174)	(0.0105)	(0.233)	(0.00497)	(0.00462)
N	33,968	33,437	16,424	33,968	17,931	9,537
	Panel D: Birth Environment					
	HC Composite in Birth Cnty.	NICU in Birth Cnty.	Cesarean Rate in Birth Cnty.			
Closed	0.0572*	0.0357***	-0.00623***			
	(0.0251)	(0.0103)	(0.00172)			
N	8,890	33,968	30,605			

Note: Estimates come from the two-way fixed effects (TWFE) specifications displayed in Figures 2–6 and 8, 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. ***, **, * indicate significance at the 0.1%, 1%, and 5% levels.

Table A3: Specification Checks (Part 1)

	(1)	(2)	(3)	(4)	(5)
Birth in Cnty. of Residence	-0.283***	-0.276***	-0.275***	-0.276***	-0.274***
	(0.00734)	(0.00743)	(0.00742)	(0.00746)	(0.00725)
Birth in Hospital	-0.00252**	-0.00114	-0.00138	-0.00178*	-0.000832
	(0.000858)	(0.000887)	(0.000860)	(0.000832)	(0.000699)
Prenatal Visits	-0.157***	-0.161***	-0.165***	-0.162***	-0.193***
	(0.0419)	(0.0427)	(0.0430)	(0.0408)	(0.0397)
Birth Weight	-3.361	-1.459	-0.856	1.089	-0.137
	(1.776)	(1.847)	(1.834)	(1.831)	(1.921)
Low Birth Wt.	-0.000160	-0.000372	-0.000552	-0.000790	-0.000538
	(0.000628)	(0.000648)	(0.000647)	(0.000654)	(0.000728)
Weeks Gestation	-0.0657***	-0.0516***	-0.0465***	-0.0273**	-0.0208*
	(0.00867)	(0.00895)	(0.00886)	(0.00876)	(0.00911)
Premature (<37 Weeks)	0.00244**	0.00160	0.00142	0.0000320	-0.000394
	(0.000861)	(0.000895)	(0.000902)	(0.000913)	(0.00103)
Gestation 37-39 Weeks	0.0171***	0.0148***	0.0135***	0.00953***	0.00833***
	(0.00191)	(0.00197)	(0.00192)	(0.00186)	(0.00195)
Induced at 37-39 Weeks	0.0210***	0.0186***	0.0171***	0.0107***	0.00898***
	(0.00203)	(0.00205)	(0.00199)	(0.00179)	(0.00184)
Induced	0.0252***	0.0226***	0.0209***	0.0133***	0.0122***
	(0.00298)	(0.00302)	(0.00299)	(0.00264)	(0.00272)
N	48,825	48,825	48,820	48,820	48,546
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 0.1%, 1%, and 5% levels.

Table A4: Specification Checks (Part 2)

	(1)	(2)	(3)	(4)	(5)
Cesarean	-0.0105***	-0.0107***	-0.0108***	-0.0123***	-0.0140***
	(0.00206)	(0.00209)	(0.00208)	(0.00200)	(0.00198)
N	48,791	48,791	48,786	48,786	48,512
Sample Years			1989-2019		
Low Apgar (<8)	0.0000495	-0.00256*	-0.00260*	-0.00247*	-0.00256*
	(0.00128)	(0.00128)	(0.00127)	(0.00124)	(0.00129)
N	48,002	48,002	47,997	47,997	47,723
Sample Years			1989-2019		
Infant Mortality Rate	0.107	0.00927	0.0165	0.0179	0.0000707
	(0.167)	(0.174)	(0.174)	(0.174)	(0.200)
N	48,825	48,825	48,820	48,820	48,546
Sample Years			1989-2019		
Neonatal Mortality Rate	0.0272	-0.0913	-0.0761	-0.0566	-0.0734
	(0.138)	(0.142)	(0.142)	(0.143)	(0.161)
N	48,825	48,825	48,820	48,820	48,546
Sample Years			1989-2019		
Infant Composite (1989-2006)	0.00723	0.00887	0.0104	0.0160*	0.0133*
	(0.00714)	(0.00725)	(0.00732)	(0.00732)	(0.00651)
N	25,209	25,209	25,204	25,204	25,063
Sample Years			1989-2006		
Maternal Composite (1989-2006)	-0.00666	-0.00600	-0.00563	-0.00284	-0.000557
	(0.00346)	(0.00372)	(0.00371)	(0.00357)	(0.00324)
N	27,720	27,720	27,715	27,715	27,559
Sample Years			1989-2006		
Maternal Composite (2009-2019)	-0.0153***	-0.0152***	-0.0148***	-0.0146***	-0.0141***
	(0.00382)	(0.00390)	(0.00393)	(0.00385)	(0.00390)
N	14,463	14,463	14,463	14,463	14,377
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 0.1%, 1%, and 5% levels.

Table A5: Specification Checks (Part 3)

	(1)	(2)	(3)	(4)	(5)
HC Composite in Birth Cnty.	0.107***	0.107***	0.105***	0.103***	0.100***
	(0.0231)	(0.0230)	(0.0228)	(0.0226)	(0.0228)
N	13,030	13,030	13,030	13,030	12,940
Sample Years			2010-2019		
NICU in Birth Cnty.	0.0356***	0.0431***	0.0423***	0.0420***	0.0392***
	(0.00980)	(0.00950)	(0.00943)	(0.00889)	(0.00831)
N	34,650	34,650	34,650	34,650	34,452
Sample Years			1995-2016		
Cesarean Rate in Birth Cnty.	-0.00991***	-0.00950***	-0.00945***	-0.00878***	-0.00865***
	(0.00178)	(0.00177)	(0.00176)	(0.00176)	(0.00163)
N	40,042	40,042	40,040	40,035	39,796
Sample Years			1992-2019		
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%, and 10% levels.

Table A6: Effects of Closures with Sample Limited to 5-Year Window Around Closure

	Panel A: Birth Location, Prenatal Visits and Birthweight					
	Birth in Cnty. of Residence	Birth in Hospital	Prenatal Visits	Birthweight	Low Bir. Wt.	
No OB Unit	-0.222***	-0.00128*	-0.166***	-5.400*	0.000841	
	(0.00609)	(0.000543)	(0.0299)	(2.487)	(0.00106)	
N	36,291	36,291	36,291	36,291	36,291	
	Panel B: Gestation and Induction					
	Weeks Gestation	Premature (<37 Weeks)	Gestation 37-39 Weeks	Induced at 37-39 Weeks	Induced Ever	
No OB Unit	-0.0275*	-0.000276	0.00964***	0.00932***	0.00835***	
	(0.0115)	(0.00133)	(0.00237)	(0.00258)	(0.00170)	
N	36,291	36,291	36,291	36,276	36,276	
	Panel C: Maternal and Infant Health Outcomes					
	Cesarean	Low APGAR	Infant Morbid. (Unrevised)	Infant Mortality Rate	Maternal Morbid. (Unrevised)	Maternal Morbid. (Revised)
No OB Unit	-0.00593**	-0.00257	0.00496	0.606	-0.00732	-0.0150**
	(0.00210)	(0.00135)	(0.00800)	(0.351)	(0.00412)	(0.00483)
N	36,277	35,814	18,969	36,291	20,720	10,598
	Panel D: Birth Environment					
	HC Composite in Birth Cnty.	NICU in Birth Cnty.	Cesarean Rate in Birth Cnty.			
No OB Unit	0.0919***	0.0356***	-0.00801***			
	(0.0202)	(0.00628)	(0.00136)			
N	9,587	25,714	31,022			

Note: For all counties experiencing a closure, samples are limited to a 5-year window around closure (i.e., two year prior to closure, the year of closure, and two years post-closure). ***, **, * indicate significance at the 0.1%, 1%, and 5% levels.

Table A7: Alternative estimates exploiting variation from all OB unit closures and openings, with no sample restrictions

	Panel A: Birth Location, Prenatal Visits and Birthweight					
	Birth in Cnty. of Residence	Birth in Hospital	Prenatal Visits	Birthweight	Low Bir. Wt.	
No OB Unit	-0.307***	-0.00184**	-0.151***	-0.818	-0.000664	
	(0.00654)	(0.000689)	(0.0348)	(1.492)	(0.000539)	
N	91,667	91,667	91,666	91,667	91,667	
	Panel B: Gestation and Induction					
	Weeks Gestation	Premature (<37 Weeks)	Gestation 37-39 Weeks	Induced at 37-39 Weeks	Induced Ever	
No OB Unit	-0.0342***	0.000585	0.00921***	0.0134***	0.0104***	
	(0.00740)	(0.000726)	(0.00159)	(0.00249)	(0.00168)	
N	91,667	91,667	91,667	91,591	91,591	
	Panel C: 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***	0.00793	0.0522	-0.00224	-0.00954**
	(0.00167)	(0.00112)	(0.00519)	(0.141)	(0.00261)	(0.00337)
N	91,589	89,473	46,444	91,667	51,842	27,590
	Panel D: Birth Environment					
	HC Composite in Birth Cnty.	NICU in Birth Cnty.	Cesarean Rate in Birth Cnty.			
No OB Unit	0.0984***	0.0475***	-0.00710***			
	(0.0214)	(0.00816)	(0.00158)			
N	24,690	64,988	76,485			

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). Furthermore, this analysis includes none of the sample restrictions described in Section 3. ***, **, * indicate significance at the 0.1%, 1%, and 5% levels.

Table A8: Heterogeneous Effects by Predicted Cesarean Need & Mode of Delivery

	Cesarean Need Tercile			Mode of Delivery	
	0-33	33-66	66-100	Vaginal	Cesarean
Cesarean Delivery	-0.005**	-0.013***	-0.009**	-	-
	(0.002)	(0.002)	(0.003)		
Difference (p-value)	-	0.000	0.083		
Mean Dep. Var.	0.078	0.177	0.601		
N	47,329	47,311	47,326		
Low Apgar (<8)	-0.002	-0.004*	-0.003	-0.002	-0.002
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
Difference (p-value)	-	0.253	0.580	-	0.825
Mean Dep. Var.	0.026	0.040	0.054	0.036	0.061
N	45,980	45,953	45,968	47,811	47,249
Infant Morbidity (U)	0.011	0.011	0.004	0.009	0.005
	(0.007)	(0.008)	(0.008)	(0.007)	(0.010)
Difference (p-value)	-	0.838	0.349	-	0.605
Mean Dep. Var.	-0.020	0.016	0.041	0.005	0.065
N	24,852	24,828	25,007	24,970	24,925
Infant Mortality Rate	0.767*	-0.665	-0.393	0.060	0.196
	(0.309)	(0.343)	(0.713)	(0.240)	(0.720)
Difference (p-value)	-	0.001	0.115	-	0.920
Mean Dep. Var.	6.07	5.15	8.40	5.85	7.55
N	37,881	37,863	37,878	39,328	39,291
Maternal Morbidity (U)	-0.009*	-0.001	-0.009	-0.004	-0.012
	(0.004)	(0.005)	(0.006)	(0.004)	(0.008)
Difference (p-value)	-	0.212	0.866	-	0.285
Mean Dep. Var.	-0.033	0.009	0.018	-0.012	0.014
N	27,388	27,370	27,532	27,552	27,458
Maternal Morbidity (R)	-0.011**	-0.015**	-0.015*	-0.019***	-0.006
	(0.004)	(0.005)	(0.007)	(0.004)	(0.008)
Difference (p-value)	-	0.505	0.627	-	0.131
Mean Dep. Var.	-0.012	0.021	0.040	0.016	0.035
N	15,876	15,871	15,878	14,440	14,189

Notes: The first three columns stratify the sample based on predicted C-section need. C-section need 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 could not be calculated for state-years using the unrevised birth certificates after 2009, and those state-years are omitted in these estimates (approximately 2.8% of the sample). The second two columns stratify the sample based on actual mode of delivery (vaginal or Cesarean). "Difference (p-value)" represents a test of equality against the first group in the category (e.g., in the "33-66" column it represents a test of the 33-66th percentile against the 0-33rd percentile). The sample sizes vary across columns because observations with no births in a given group-county-year are dropped. For infant mortality, these analyses require using the linked birth-infant death files which are only available for 1989-1991 and 1996-2017 whereas other analyses of infant mortality use the unlinked mortality files available for 1989-2019. ***, **, * indicate significance at the 0.1%, 1%, and 5% levels.

Table A9: Heterogeneous Effects of Closures by Race, Education, and Age: Outcomes Part 1

	Race			Education		Age		
	White NH	Hispanic	Black NH	No College	Some College+	Age<25	Age 25-34	Age 35+
Birthweight	-0.90	-3.34	-8.17	0.74	-4.50	0.44	-2.65	1.76
	(2.13)	(7.73)	(10.25)	(2.89)	(2.83)	(2.58)	(2.22)	(4.97)
Difference (p-value)	-	0.821	0.465	-	0.127	-	0.274	0.746
Mean Dep. Var.	3335	3288	3057	3255	3364	3245	3347	3325
N	48,796	44,643	38,611	48,282	48,365	48,819	48,820	48,715
Weeks Gestation	-0.048***	-0.088*	-0.093*	-0.043***	-0.052***	-0.049***	-0.046***	-0.040*
	(0.010)	(0.036)	(0.046)	(0.013)	(0.012)	(0.012)	(0.010)	(0.019)
Difference (p-value)	-	0.274	0.363	-	0.504	-	0.913	0.663
Mean Dep. Var.	38.8	38.7	38.1	38.7	38.8	38.8	38.8	38.4
N	48,796	44,631	38,607	48,283	48,365	48,819	48,820	48,715
Premature (<37 Weeks)	0.0016	0.0005	0.0099	0.0019	0.0024	0.0025*	0.0010	-0.0001
	(0.0010)	(0.0042)	(0.0054)	(0.0014)	(0.0013)	(0.0013)	(0.0011)	(0.0027)
Difference (p-value)	-	0.767	0.146	-	0.754	-	0.418	0.418
Mean Dep. Var.	0.111	0.121	0.192	0.131	0.106	0.126	0.111	0.144
N	48,796	44,631	38,607	48,283	48,365	48,819	48,820	48,715
Gestation 37-39 Weeks	0.006**	0.011	-0.009	0.003	0.005*	0.003	0.005**	0.008*
	(0.002)	(0.006)	(0.007)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Difference (p-value)	-	0.482	0.032	-	0.417	-	0.136	0.136
Mean Dep. Var.	0.243	0.258	0.274	0.242	0.255	0.233	0.254	0.283
N	48,796	44,631	38,607	48,283	48,365	48,819	48,820	48,715
Induction at 37-39 Weeks	0.006***	0.004	-0.001	0.007***	0.004**	0.006***	0.006***	0.010***
	(0.001)	(0.003)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Difference (p-value)	-	0.498	0.094	-	0.022	-	0.523	0.071
Mean Dep. Var.	0.058	0.048	0.049	0.050	0.064	0.053	0.058	0.060
N	48,778	44,581	38,508	48,236	48,295	48,757	48,767	48,594

Notes: Each coefficient represents a separate regression. "Difference (p-value)" represents a test of equality against the first group in the category (e.g., in Column 2 it represents a test of Hispanic against White non-Hispanic). The sample sizes vary across columns because observations with no births in a given group-county-year are dropped. Educational attainment was not measured from 2009-2013 for a small number of states that had not yet switched to revised birth certificates; those observations are dropped. ***, **, * indicate significance at the 0.1%, 1%, and 5% levels.

Table A10: Heterogeneous Effects of Closures by Race, Education, and Age: Outcomes Part 2

	Race			Education		Age		
	White NH	Hispanic	Black NH	No College	Some College+	Age<25	Age 25-34	Age 35+
Cesarean Delivery	-0.009***	-0.028***	-0.011	-0.013***	-0.009***	-0.012***	-0.008***	-0.009*
	(0.002)	(0.007)	(0.008)	(0.003)	(0.002)	(0.003)	(0.002)	(0.004)
Difference (p-value)	-	0.008	0.759	-	0.128	-	0.074	0.492
Mean Dep. Var.	0.292	0.276	0.319	0.275	0.313	0.249	0.315	0.383
N	48,778	44,578	38,513	48,237	48,293	48,757	48,767	48,593
Low Apgar (<8)	-0.002	-0.005	-0.002	-0.005**	-0.001	-0.003*	-0.002	-0.003
	(0.001)	(0.003)	(0.004)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Difference (p-value)	-	0.317	0.984	-	0.013	-	0.340	0.667
Mean Dep. Var.	0.039	0.034	0.049	0.043	0.038	0.042	0.038	0.049
N	47,736	42,943	36,813	47,188	47,069	47,644	47,580	46,887
Infant Composite (Unrevised)	0.008	-0.028	0.021	0.006	0.014*	0.012	0.008	-0.001
	(0.007)	(0.019)	(0.023)	(0.008)	(0.007)	(0.008)	(0.007)	(0.012)
Difference (p-value)	-	0.050	0.514	-	0.290	-	0.552	0.248
Mean Dep. Var.	0.009	0.032	0.007	0.016	0.005	0.016	0.005	0.029
N	25,230	21,383	18,157	24,844	24,971	24,852	25,078	24,567
Infant Mortality Rate	0.280	1.078	-1.805	-0.480	-0.273	-0.127	0.003	0.599
	(0.257)	(0.939)	(1.959)	(0.363)	(0.362)	(0.319)	(0.271)	(0.771)
Difference (p-value)	-	0.492	0.330	-	0.664	-	0.761	0.425
Mean Dep. Var.	6.505	5.207	9.872	7.983	5.093	7.903	5.831	6.738
N	37,773	34,727	30,015	37,260	37,344	37,797	37,797	37,707

Notes: Each coefficient represents a separate regression. "Difference (p-value)" represents a test of equality against the first group in the category (e.g., in Column 2 it represents a test of Hispanic against White non-Hispanic). The sample sizes vary across columns because observations with no births in a given group-county-year are dropped. Educational attainment was not measured from 2009-2013 for a small number of states that had not yet switched to revised birth certificates; those observations are dropped. In these analyses, the infant mortality data are derived from the linked birth-infant death files and are only available for 1989-1991 and 1996-2017 whereas other analyses of infant mortality use the unlinked mortality files available for 1989-2019. The linked data are required for this analysis because data on mother's demographics are required. ***, **, * indicate significance at the 0.1%, 1%, and 5% levels.

Table A11: Heterogeneous Effects of Closures by Race, Education, and Age: Outcomes Part 3

	Race			Education		Age		
	White NH	Hispanic	Black NH	No College	Some College+	Age<25	Age 25-34	Age 35+
Maternal Composite (Unrevised)	-0.005	0.010	-0.001	-0.005	-0.005	-0.005	-0.007	0.012
	(0.004)	(0.018)	(0.024)	(0.004)	(0.004)	(0.005)	(0.004)	(0.008)
Difference (*p*-value)	-	0.404	0.715	-	0.893	-	0.638	0.084
Mean Dep. Var.	-0.005	0.019	-0.013	-0.003	-0.001	0.007	-0.012	-0.014
N	27,740	23,847	19,981	27,381	27,500	27,390	27,597	27,123
Maternal Composite (Revised)	-0.013**	-0.042*	-0.022	-0.018***	-0.006	-0.020***	-0.009*	-0.021
	(0.004)	(0.018)	(0.016)	(0.005)	(0.006)	(0.005)	(0.005)	(0.013)
Difference (*p*-value)	-	0.129	0.676	-	0.077	-	0.102	0.834
Mean Dep. Var.	0.015	0.026	0.038	0.020	0.015	0.016	0.018	0.053
N	14,095	13,892	12,374	16,757	16,840	14,775	14,065	13,572

Notes: Each coefficient represents a separate regression. "Difference (*p*-value)" represents a test of equality against the first group in the category (e.g., in Column 2 it represents a test of Hispanic against White non-Hispanic). The sample sizes vary across columns because observations with no births in a given group-county-year are dropped. Educational attainment was not measured from 2009-2013 for a small number of states that had not yet switched to revised birth certificates; those observations are dropped. ***, **, * indicate significance at the 0.1%, 1%, and 5% levels.

Table A12: 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 A13: Closure Probit 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%, and 10% levels. Estimates are from a cross-sectional probit 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

As an alternative to our closure measure from NVSS data, we can use data from the AHA Annual Surveys for 1995-2016 to identify obstetric unit 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:[32]

1. Set $OB_{Open} = 0$ if the hospital reports zero obstetric beds, zero bassinets, and < 10 births (22,950 hospital-years).
2. Set $OB_{Open} = 1$ if the hospital reports > 0 obstetric beds, > 0 bassinets, and > 10 births (58,964 hospital-years).
3. If OB_{Open} is still not defined, set $OB_{Open} = 1$ if the hospital reports > 25 births (15,457 hospital-years).
4. If OB_{Open} is still not defined, set $OB_{Open} = 0$ if the hospital reports < 5 births (9,434 hospital-years).
5. If OB_{Open} is still not defined, set $OB_{Open} = 1$ if the hospital reports > 0 bassinets (798 hospital-years).
6. If OB_{Open} is still not defined, set $OB_{Open} = 0$ if the hospital reports zero bassinets (502 hospital-years).

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 A2. The results are qualitatively similar across all outcomes.

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

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.

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 US 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.[33]

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.[34] 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 A12 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.

[33] 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.

[34] 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.

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 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 and are 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 NICU bed (17,836 hospital-years).
2. For hospital-years with non-missing data, assign NICU=0 for those with zero NICU beds (71,606 hospital-years).
3. For hospital-years with missing data, assign NICU=0 if NICU=0 for the hospital in every other year (14,080 hospital-years).
4. For hospital-years with missing data, assign NICU=1 if NICU=1 for the hospital in every other year (1,336 hospital-years).
5. For hospital-years with missing data, assign NICU=0 if the hospital has no non-missing values for any year (778 hospital-years).

6. For hospital-years with missing data, assign NICU equal to the hospital's most recent non-missing value (2,263 hospital-years).
7. For hospital-years with missing data, assign NICU equal to the hospital's closest future non-missing value (216 hospital-years).

A.3 Details of the Econometric Approach

A.3.1 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 A3 to A5. 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 (US Census Bureau, 2010). 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 A13.

A.3.2 Event Study Specification

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

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.3.3 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 7,348 ATTs, where 711 (9.6%) 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 -0.015 (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.

A.3.4 Sample Restrictions for C-Section Mechanism Analysis

This section refers to the estimates presented in Figure 8. This analysis requires restricting the sample in three ways.

1. The first three years (1989-1992) of the overall sample are dropped to account for the fact

that the outcome in Figure 8A and the C-section gaps in Figure 8B 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 counties 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.