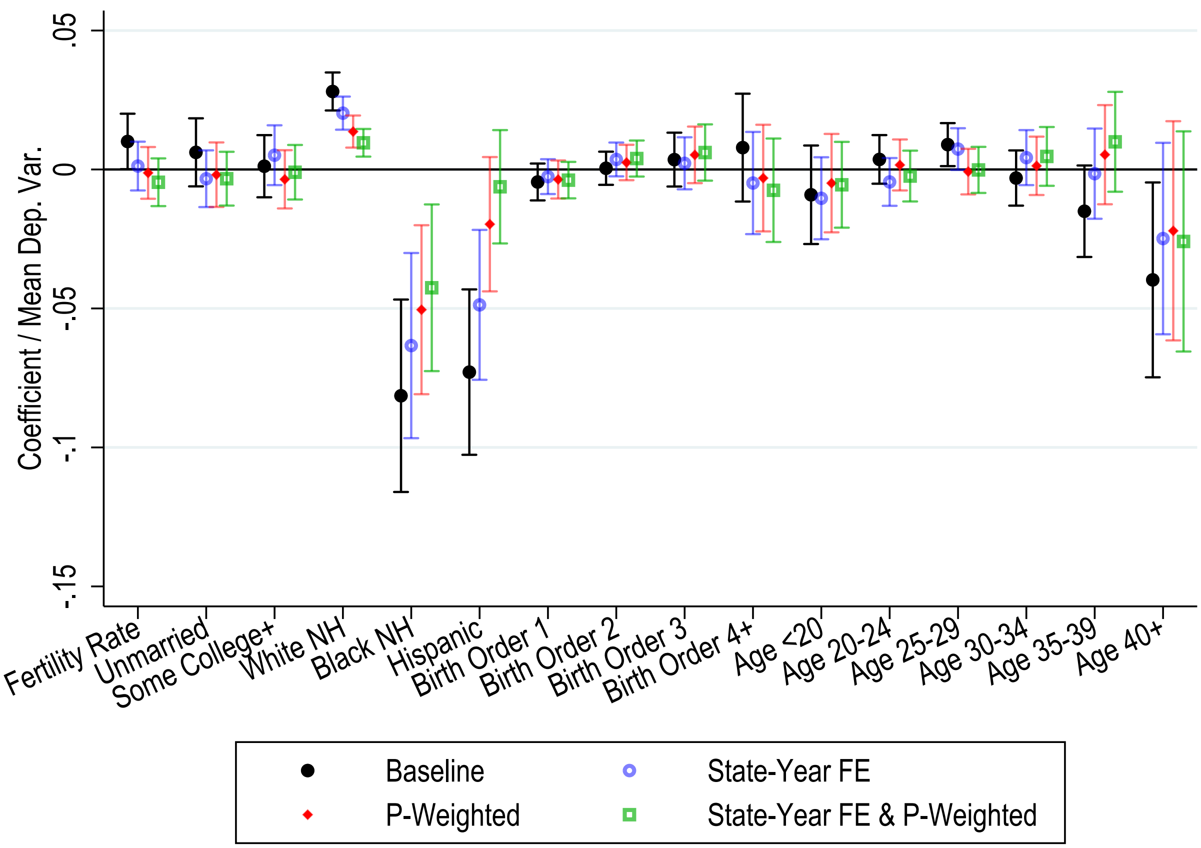
ONLINE APPENDIX

Fischer, Royer, & White

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

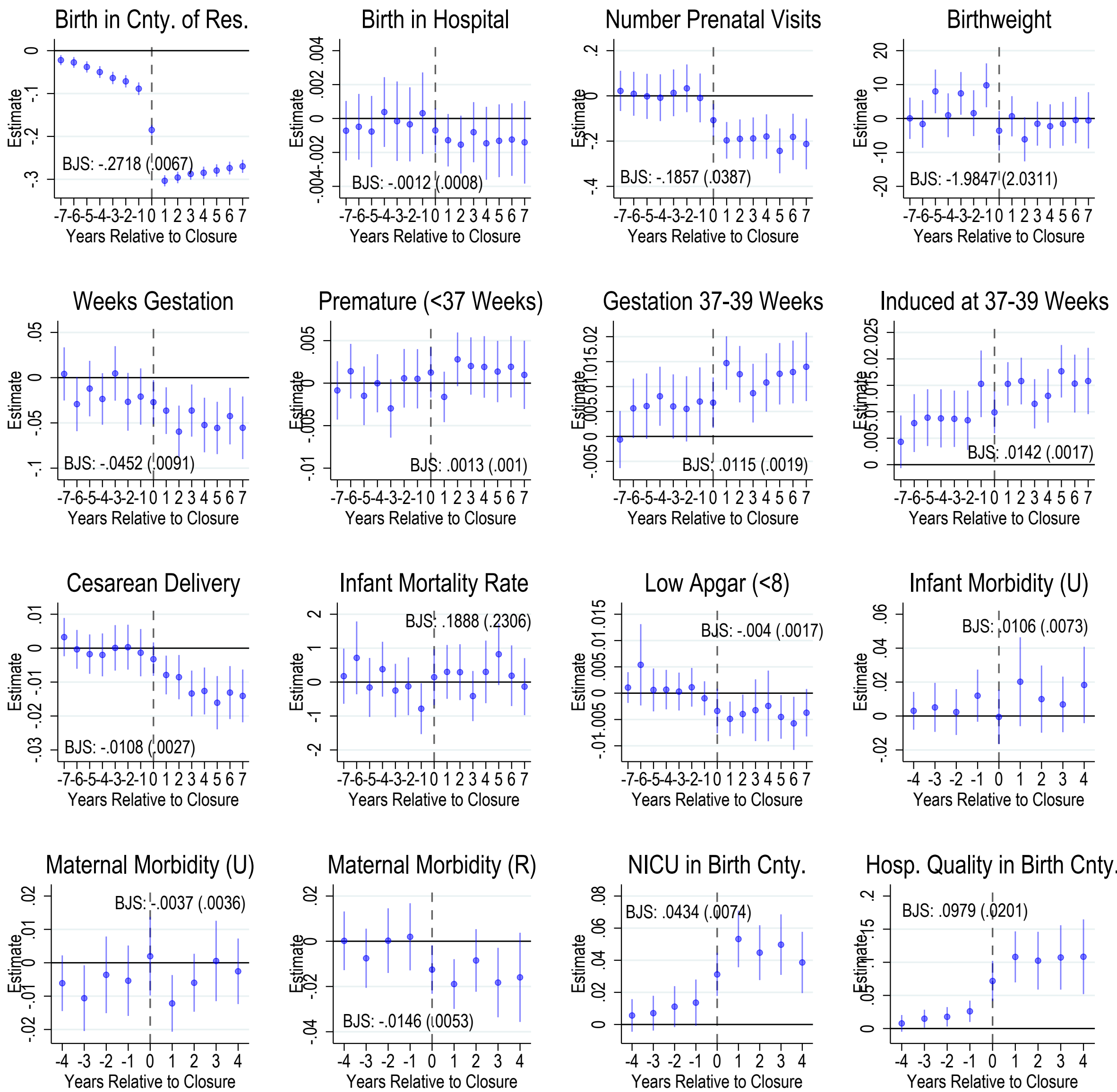
Appendix Figures

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



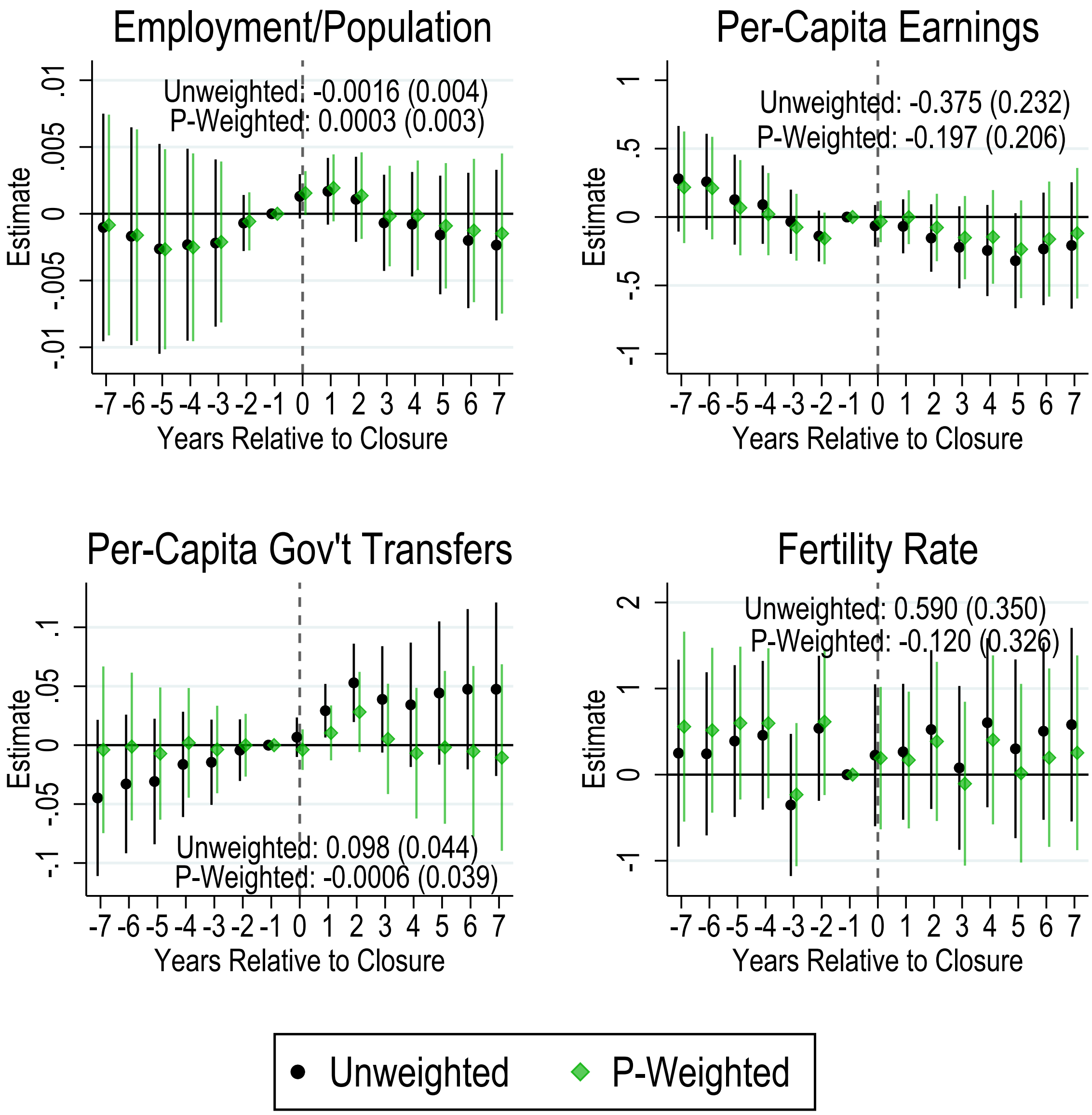
Notes: For comparability across outcomes, all coefficient estimates are divided by the mean of the dependent variable. The baseline specification (black solid circles) is described in Eq. (1). The second specification (blue open circles) adds state-by-year fixed effects. The third specification (red 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



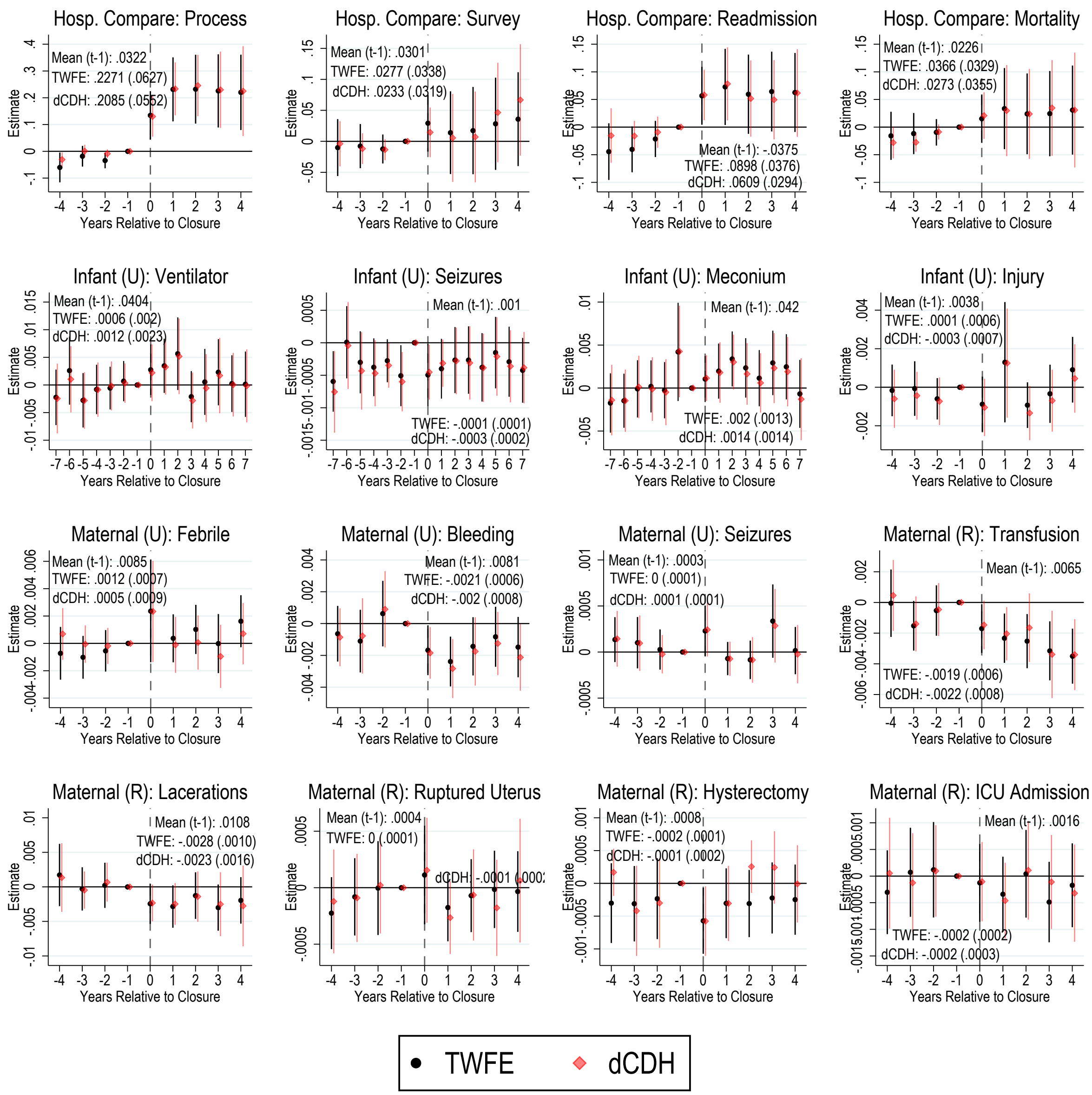
Notes: These plots replicate the 16 estimates presented in Figures 2–5 and 8 using the Borusyak et al. (2021) imputation-based difference-in-differences estimator. The point estimate labeled "BJS" on each plot represents the average effect across the post-treatment periods $t = 0$ through $t = 7$. All 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



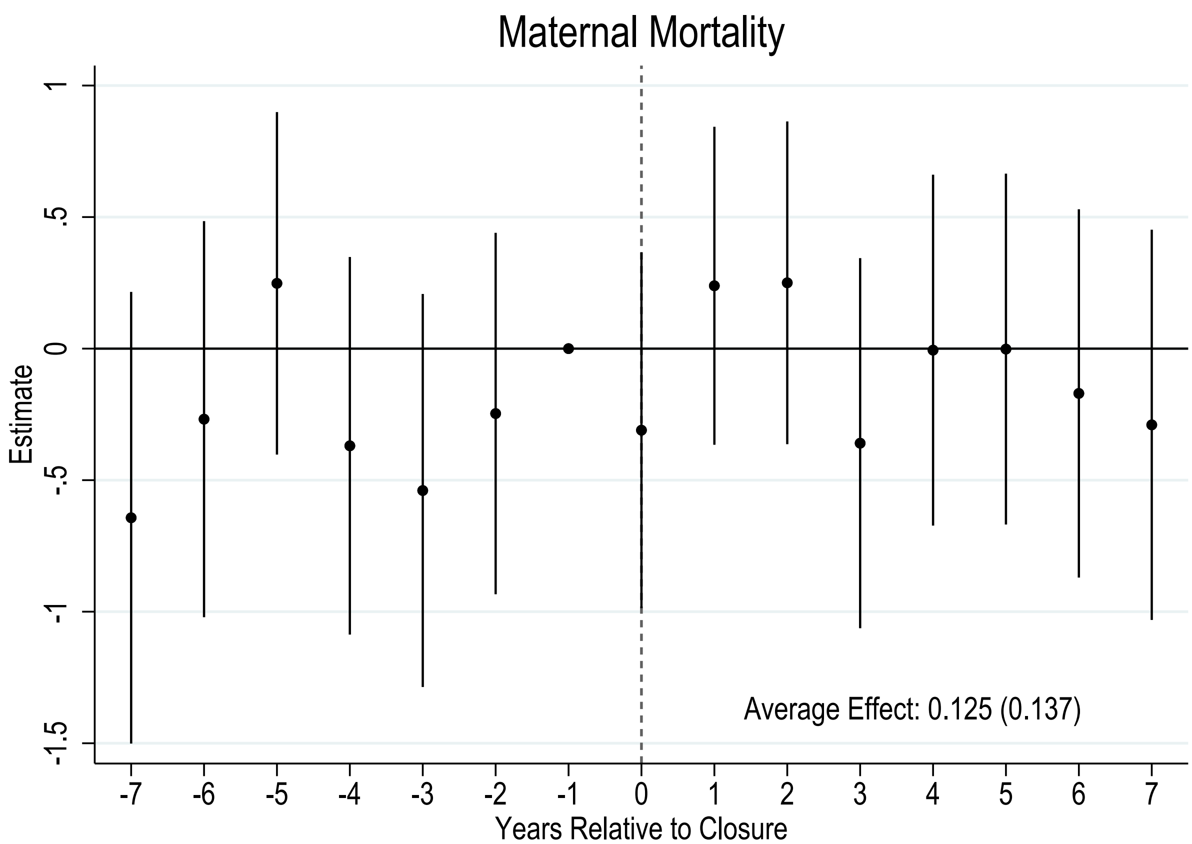
Notes: See Figure 2 for general notes on interpreting the event studies. “Unweighted” refers to our main specification, and “P-Weighted” refers to a specification in which counties are weighted by the propensity to experience a closure. The process of calculating propensity score weights is described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

Figure A4: Effect of Closures on Components of Composite Measures



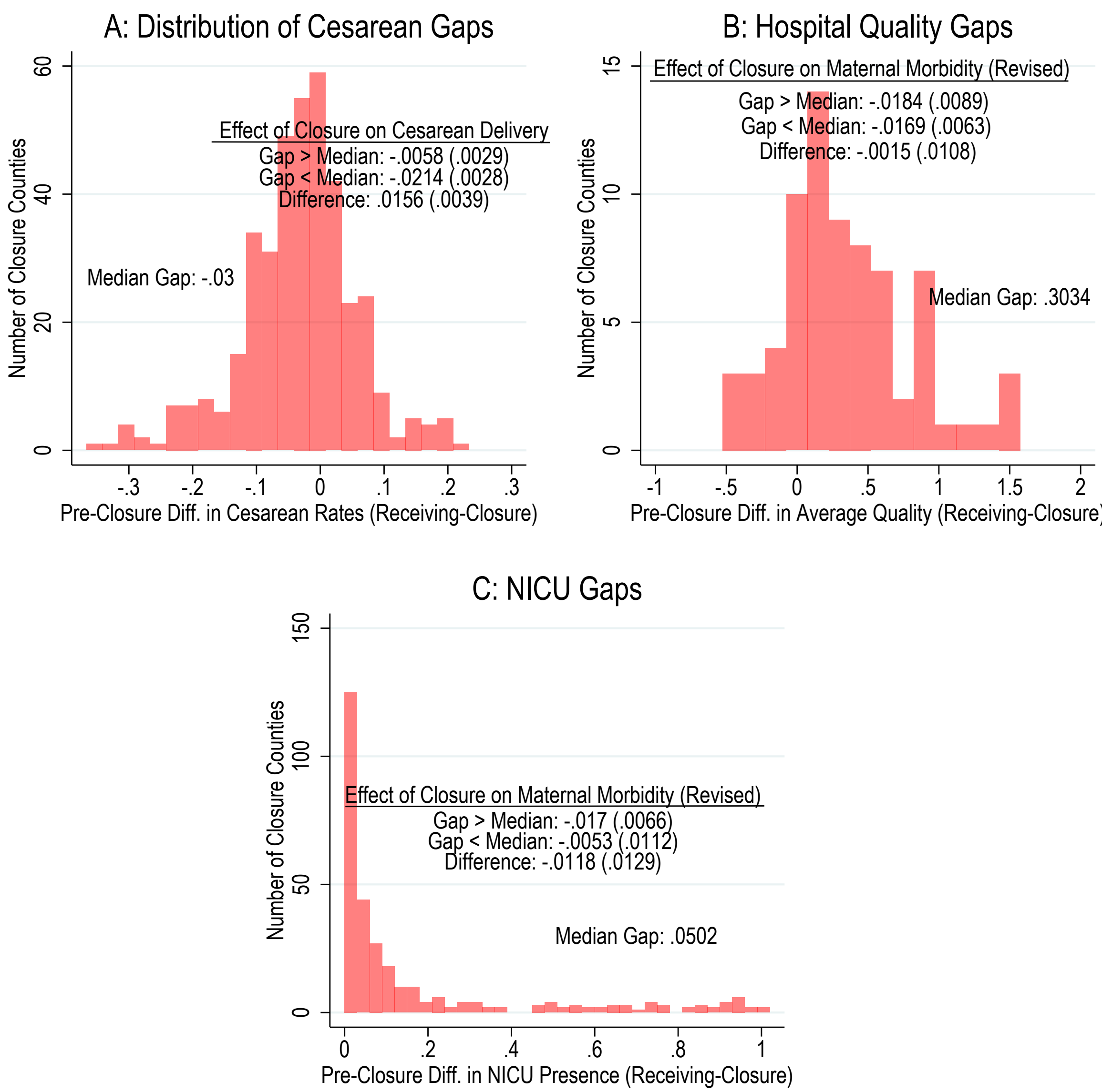
Notes: See Figure 2 for general notes on interpreting event studies. The top four plots show effects of closures on the four components of the hospital quality composite from Hospital Compare. We follow Doyle et al. (2019) in constructing the four measures: process measures, patient survey measures, 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

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

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

A.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.