Online Appendix: Unintended Consequences of Policy Interventions: Evidence from Mandated Health Insurance Coverage for IVF Treatment, Arezou Zaresani and Lucie Schmidt, American Journal of Health Economics, Volume 10, Issue 3, 2024, https://doi.org/10.1086/725908

Appendix

A Summary statistics

Table A.1: Summary statistics for Society for Assisted Reproductive Technology (SART) infertility clinic data, 1996-2010

	Never mandated states (control group)		*Mandate to cover states (treatment group)*	
	1996-2004	2005-2010	1996-2004	2005-2010
Total number of cycles	231,699	201,147	199,085	176,306
Average number of embryos transferred for all women	3.25 (0.01)	2.45 (0.01)	3.18 (0.03)	2.47 (0.02)
Multiple births per hundred live births for all women	34.87 (0.36)	30.95 (0.42)	33.48 (0.45)	28.92 (0.50)
Cycles for women 35 and older (%)	48.10	50.13	56.17	59.40
Average number of embryos transferred for women 35 and older	3.39 (0.02)	2.64 (0.01)	3.30 (0.02)	2.66 (0.02)
Average number of embryos transferred for women under 35 years	3.12 (0.02)	2.23 (0.01)	3.03 (0.03)	2.22 (0.02)
Total number of IVF clinics	326	255	118	94

Notes: Standard deviations appear in parentheses.

Table A.2: Summary statistics for National Data Archive on Child Abuse and Neglect (NDACAN) adoption data, 1994-2015

	Never mandated states (control group)		*Mandate to cover states (treatment group)*	
	1994-2004	2005-2015	1994-2004	2005-2015
Number of adopted children per ten thousand newborn infants	5.23	8.78	6.37	6.86
Number of adopted children	103,327	188,072	37,926	36,811
Number of newborn infants	19,736,577	21,411,844	5,955,365	5,362,502
Adopting women 35 and older (%)	79.31	79.99	85.21	82.95
	(0.13)	(0.09)	(0.18)	(0.19)
Mean age of adopting mothers	40.99	41.41	42.70	42.04
	(0.02)	(0.02)	(0.04)	(0.04)
Mean age of adopting fathers	43.04	43.55	45.27	44.52
	(0.02)	(0.01)	(0.03)	(0.03)
White adopting mothers (%)	62.20	69.47	38.90	53.76
	(0.15)	(0.10)	(0.25)	(0.25)
White adopting fathers (%)	55.05	59.71	32.26	44.53
	(0.15)	(0.11)	(0.24)	(0.25)
Mean age of adopted children	3.31	3.02	3.61	3.07
	(0.01)	(0.00)	(0.01)	(0.01)
White adopted children (%)	48.93	51.76	28.47	39.31
	(0.15)	(0.11)	(0.23)	(0.25)
Adopted boys (%)	50.89	51.52	50.79	51.49
	(0.15)	(0.11)	(0.26)	(0.25)

Note: Data include children age 0-6 adopted in the US. Standard deviations appear in parentheses.

Table A.3: Summary statistics for Current Population Survey Annual Social and Economic Supplement (CPS)

	Never mandated states (control group)				*Mandate to cover states (treatment group)*			
	1975-1984	1985-1994	1995-2004	2005-2014	1975-1984	1985-1994	1995-2004	2005-2014
Women of child bearing age (18-49 years) (%)	38.66	39.38	38.12	34.78	38.34	39.53	38.20	35.11
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female labor force participation rate (%)	61.45	69.31	73.03	71.21	62.54	70.45	74.16	72.08
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Employee in firms of +500 employee (%)	16.19	16.29	15.86	13.88	17.61	17.72	17.07	14.91
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Private health insurance (%)	78.23	76.24	74.80	69.55	81.72	79.94	76.89	73.74
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Real average per capita income (2007 USD)	25,076	29,958	35,033	36,093	26,400	31,624	37,576	39,161
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Note: The sample includes working age individuals (18 to 64 years). Standard deviations appear in parentheses.

B Estimation procedure of a GSC model

Xu (2017) provides a procedure for estimating a Generalized Synthetic Control (GSC) model specified in Equation (2) as:

$$y_{it} = \delta_{it} D_{it} + X'_{it}\beta + \lambda'_i f_t + \epsilon_{it}. \tag{B.1}$$

The procedure consists of three main steps. The first step includes estimating an interactive fixed-effect model using the data only from the control group (i.e., setting $D_{it} = 0$ in Equation (B.1)). Assume that f_t and λ_i are r-vectors where r denotes the number of factors. Also assume that $F = [f_1, f_2, ..., f_T]$ and $\Lambda_{control} = [\lambda_1, \lambda_2, ..., \lambda_{control}]$ where *control* denotes the number of states in the control group and T denotes the time periods in the analysis. To identify β, F and $\Lambda_{control}$ however more constraints are required. Two constraints are imposed. First, all factors are normalized, $\frac{\widehat{F}'\widehat{F}}{|T|} = I_r$, where I_r denotes the identity matrix. Second, loadings are orthogonal to each other, $\widehat{\Lambda}'_{control}\widehat{\Lambda}_{control} = 0$. To obtain the estimated $\widehat{\beta}$, $\widehat{F}$ and $\widehat{\Lambda}_{control}$ then:

$$(\widehat{\beta}, \widehat{F}, \widehat{\Lambda}_{control}) = \underset{\widehat{\beta}, \widehat{F}, \widehat{\Lambda}_{control}}{\arg\max} \sum_{i \in control} (Y_i - X_i\widehat{\beta} - \widehat{F}\widehat{\lambda}_i)'(Y_i - X_i\widehat{\beta} - \widehat{F}\widehat{\lambda}_i), \tag{B.2}$$

$$\text{s.t.} \ \frac{\widehat{F}'\widehat{F}}{|T|} = I_r \ \text{and} \ \widehat{\Lambda}'_{control}\widehat{\Lambda}_{control} = 0.$$

The number of factors r is unknown and is estimated through a cross validation process that minimizes the prediction error of the model. The estimation process starts with a given r to obtain the corresponding $\widehat{\beta}$, $\widehat{F}$ and $\widehat{\Lambda}_{control}$. For each pre-treatment period $s \in \{1, 2, ..., T_0\}$ (T_0 denotes the number of pre-treatment periods), we hold back data of all treated states at time s. We then run an OLS regression using the rest of the pre-treatment data to obtain factor loadings for each treated unit i, $\widehat{\lambda}_{i,-s}$. We next predict the treated outcome at time s as $\widehat{y}_{is}(0) = X'_{is}\widehat{\beta} + \widehat{\lambda}_{i,-s}\widehat{f}_s$.[1]

We define the prediction error as $e_{is} = y_{is}(0) - \widehat{y}_{is}(0)$. The Mean Square Prediction

[1] $y_{it}(1)$ and $y_{it}(0)$ denote the potential outcomes for state i at time t when respectively $D_{it} = 1$ (treated) and $D_{it} = 0$ (not treated).

Error (MSPE) for a given r is defined as:

$$MSPE(r) = \sum_{s=1}^{T_0} \sum_{i \in T} \frac{e_{is}^2}{T_0} \tag{B.3}$$

This process is repeated for different values of r (we try $r \in \{1, 2, ..., 5\}$). Then, r^* corresponding to the smallest prediction error is chosen.

The factor loadings for the treated states are estimated in the second step. This is done by minimizing the MSPE of the predicted treated outcome in pretreatment periods:

$$\widehat{\lambda}_i = \arg\max_{\widehat{\lambda}_i} (Y_i^0 - X_i^0 \widehat{\beta} - \widehat{F}^0 \widehat{\lambda}_i)'(Y_i^0 - X_i^0 \widehat{\beta} - \widehat{F}^0 \widehat{\lambda}_i) \tag{B.4}$$

where "0" superscripts denote the pre-treatment time periods and $\widehat{\beta}$ and $\widehat{F}^0$ are estimated from the first step.

Finally, the third step estimates the treated counterfactual based on $\widehat{\beta}$, $\widehat{F}$ and $\widehat{\lambda}_i$. That is:

$$\widehat{y}_{it}(0) = X_{it}' \widehat{\beta} + \widehat{\lambda}_i' \widehat{f}_i \quad \text{for } i \in Treated,\ t > T_0 \tag{B.5}$$

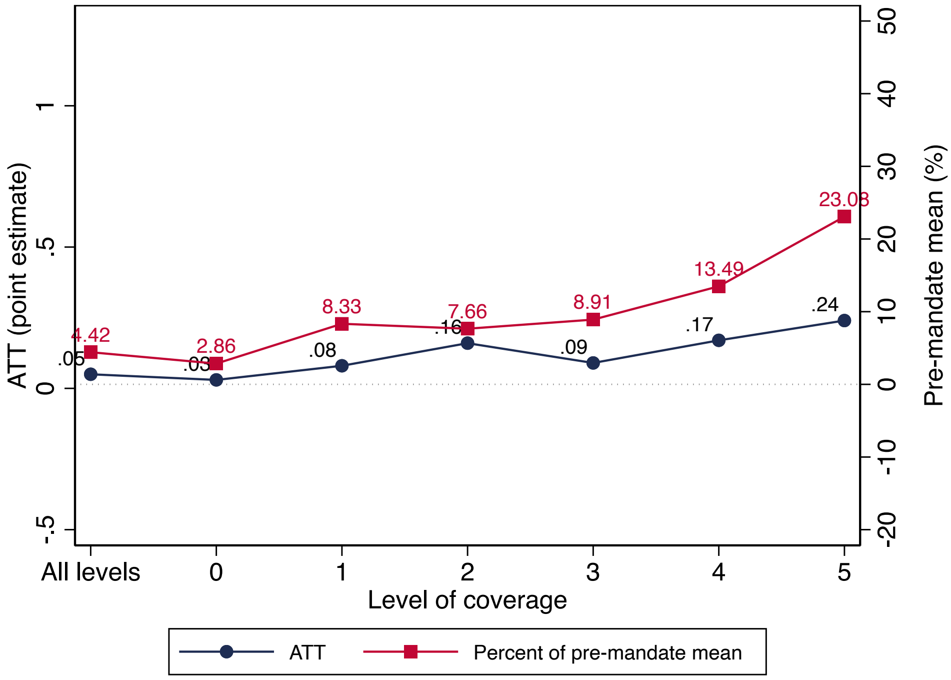
The estimated Average Treatment effect on Treated at time t, ATT_t then is:

$$\widehat{ATT}_t = \frac{1}{|Treated|} \sum_{i \in Treated} [y_{it}(1) - \widehat{y}_{it}(0)] \quad \text{for } t > T_0 \tag{B.6}$$

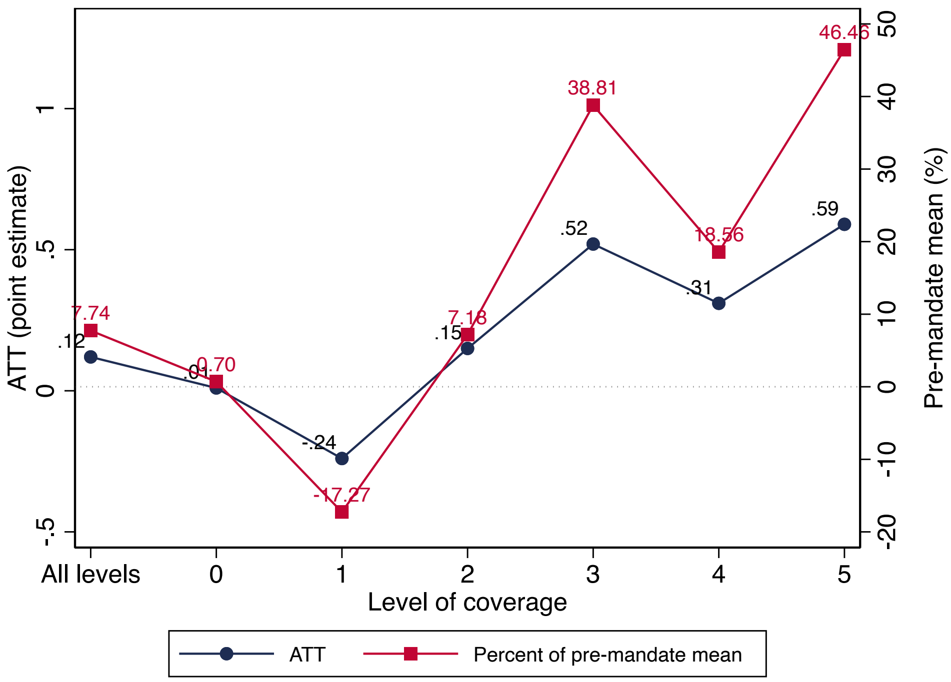
C Graphical presentation and statistical significance tests of GSC estimates

Figure C.1: GSC estimates for multiple births per hundred live births

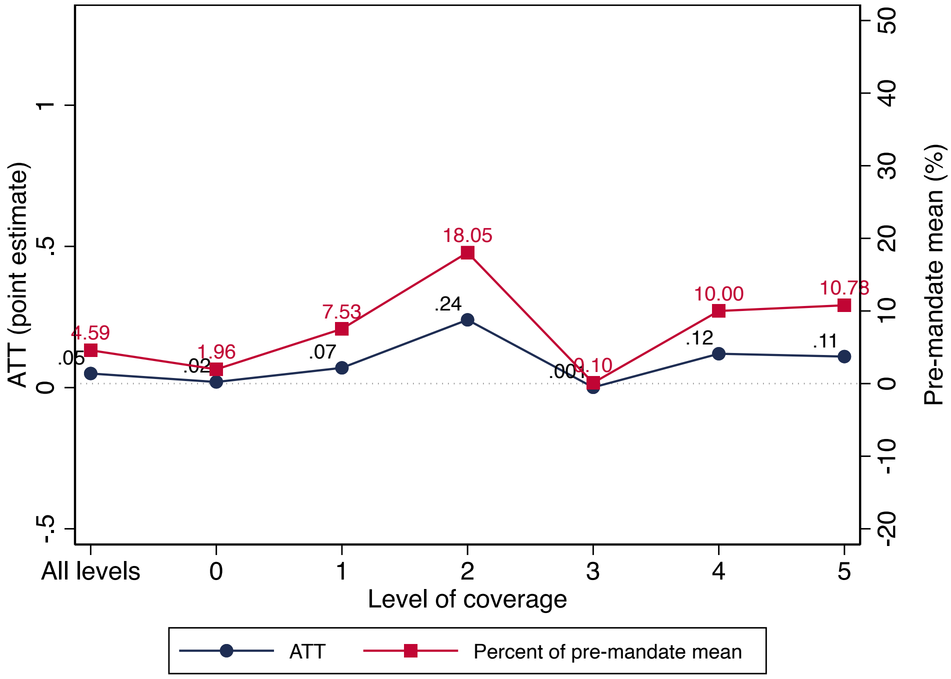
(a) All women



(b) Women 35 years and older



(c) Women under 35 years old



Note: This figure plots the estimated effects from the generosity level of coverage on the incidence of multiple births per hundred live births using the GSC model specified in Equation (2) presented in Table 4. All estimates include covariates specified in notes to Table 4.

Table C.1: Statistical significance tests of GSC estimates of the effects of generosity level of IVF coverage

(a) Number of multiple births per hundred live births

	All women	Women 35 and older	Women under 35
Level 0 vs Level 1	0.00	1.00	0.00
Level 1 vs Level 2	0.00	0.00	0.00
Level 2 vs Level 3	1.00	0.00	1.00
Level 3 vs Level 4	0.00	1.00	0.00
Level 4 vs Level 5	0.00	0.00	1.00

(b) Number of infants per thousand live births

	All women	Women 35 and older	Women under 35
Level 0 vs Level 1	0.00	1.00	0.00
Level 1 vs Level 2	0.00	0.00	0.00
Level 2 vs Level 3	1.00	0.00	1.00
Level 3 vs Level 4	0.00	0.90	0.00
Level 4 vs Level 5	0.00	0.00	1.00

Note: This table presents the 95% p-values of the two-sample Welch statistic testing $H_0 : \delta_{Level_i} = \delta_{Level_{i+1}}$ versus $H_1 : \delta_{Level_i} < \delta_{Level_{i+1}}$ for estimates with covariates presented in Table 4 and Table 5, respectively. The tests assume that the population distributions are normal, but have unequal variances.

D Other dimensions of generosity and state by state GSC estimates

Table D.1: Effects of IVF coverage on multiple births, stratified by other dimensions of mandated coverage

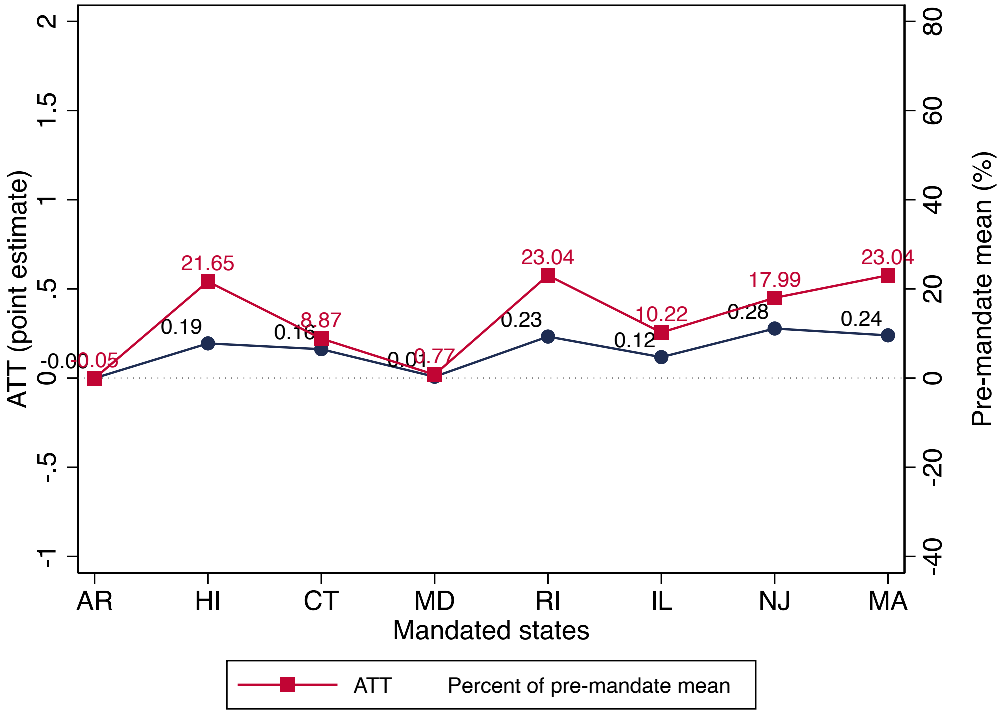
Dimension	With dimension	Without dimension	p-value of statistical significance test
	(1)	(2)	(3)
Long min infertility time	0.11	0.19***	0.00
	(0.12)	(0.07)	
Religious exemption	0.17**	0.10*	1.00
	(0.08)	(0.05)	
Firm size exemption	0.29***	0.07	1.00
	(0.09)	(0.07)	
Marital status restriction	0.08	0.21***	0.00
	(0.04)	(0.07)	
Age restriction	0.21***	0.08*	1.00
	(0.09)	(0.04)	
Restricted embryo numbers	0.16	0.27***	0.00
	(0.13)	(0.09)	
Lifetime cap	-0.02	0.27***	0.00
	(0.80)	(0.08)	

Note: This table presents the GSC estimated effects of IVF coverage (the presence of any mandate) on multiple births per hundred live births, stratified by the other dimensions of mandated coverage presented in Table 1. Each coefficient is from a separate regression. Column 1 presents the estimates from comparing *mandate to cover* states whose mandate includes a specific dimension to never mandated states. Column 2 presents the estimates from comparing *mandate to cover* states whose mandate does not include a specific dimension to never mandated states. Column 3 presents the 95% p-value of the two-sample Welch statistic testing $H_0 : \delta_{with} = \delta_{without}$ against $H_1 : \delta_{with} < \delta_{without}$. The tests assume that the population distributions are normal, but have unequal variances. The significant p-values denote the dimensions of mandated coverage that might affect the incidence of multiple births. For more details on the GSC estimates, see notes to Table 4.

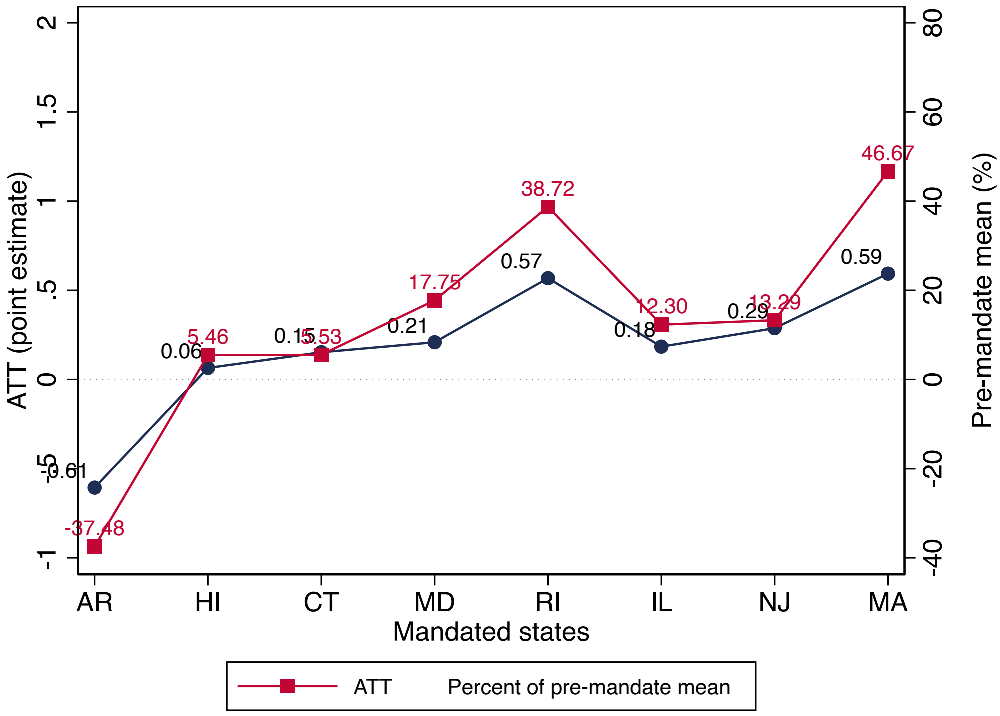
$*p < 0.10, **p < 0.05, ***p < 0.01$

Figure D.1: GSC estimates for multiple births per hundred live births

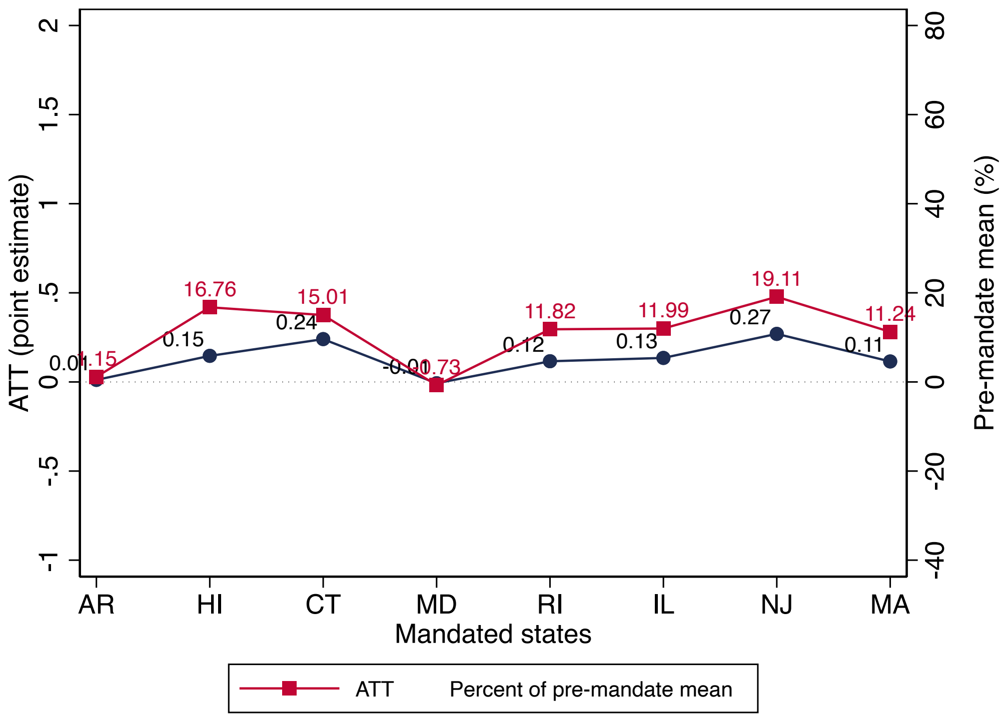
(a) All women



(b) Women 35 years and older



(c) Women under 35 years old



Note: This figure plots the estimated effects from the generosity level of coverage on the incidence of multiple births per hundred live births using the GSC model specified in Equation (2). We compare each mandated state with never mandated states. All estimates include state and time fixed effects and covariates specified in notes to Table 4.

E DD and DDD estimates

To investigate robustness of our findings from the GSC framework, we estimate the effects from mandated IVF coverage on incidence of multiple births using DD and DDD frameworks. We estimate an equation of this form for our DD model:

$$y_{it} = \alpha_0 + \alpha_1(Level_{it} \times Post_t) + \alpha_2 Level_{it} + \lambda_i + \lambda_t + \epsilon_{it} \tag{E.1}$$

where i and t denote state and time, respectively. y_{it} denotes the outcome variables: the multiple birth rate per hundred live births and the number of infants per thousand live births. $Level_{it}$ includes indicators that denote the generosity level of the mandated coverage. It is set to zero for the never-mandated states. $Post_{it}$ is a dummy variable switching on two years after the mandated coverage is enacted. It is set to zero for never-mandated states. The vector X_{it} includes the same set of state-level time-varying covariates used in the GSC analysis. λ_i and λ_t are state and time fixed effects. ϵ_{it} captures any remaining unobserved factors affecting the outcome variable. The coefficient of interest is α_1, which captures the effect of mandated coverage's generosity on the incidence of multiple births.

We estimate the following equation in our DDD model:

$$\begin{aligned} y_{ita} =& \alpha_0 + \alpha_1(Level_{it} \times Plus35_a \times Post_{it}) + \alpha_2(Level_{it} \times Plus35_a) \\ &+ \alpha_3(Post_{it} \times Plus35_a) + \alpha_4(Level_{it} \times Post_{it}) + \alpha_5 X'_{ita} \\ &+ \lambda_i + \lambda_t + \lambda_a + \epsilon_{ita} \end{aligned} \tag{E.2}$$

where a denotes women's age. $Plus35_a$ is a dummy indicating women 35 years and older. λ_a is the age fixed effects. The coefficient of interest is α_1 which captures the effect of the number of covered cycles on mothers of 35 years and older in mandated states relative to mothers younger than 35 years.

We aggregate the birth data into state-year and state-year-age cells for estimating the DD and DDD models, respectively. The estimation results are presented in Table E.1 and Table **??**. The estimates in each table's first and second columns show the replicated estimates from Buckles (2013) including the states with mandates in the 2000s in our treatment group and using a longer pre-mandate period.[2] Our estimates are much larger in magnitude and are statistically significant. Our estimated effects from any mandate on multiple birth rate and the number of infants per thousand live births are respectively 0.10 (p-value < 0.001) and 1.07 (p-value < 0.001) versus 0.02 (p-value > 0.10) and 0.28 (p-value > 0.10).

Overall the estimated effects from DDD and DD models confirm findings from our GSC framework, although the estimated effects are relatively larger than the GSC estimates. These findings suggest that more generous coverage is associated with an increase in the incidence of multiple births. The estimated effects are larger for older women than those for younger women.

Figure E.1 plots the estimated weight of each mandated state over the years, which are the residuals from a regression of a mandated coverage indicator on state and year fixed effects, scaled by the sum of the squared residuals across a pooled sample of mandated and never mandated states (see de Chaisemartin and D'Haultfoeuille (2020) for more details). None of the treated states has a negative weight, suggesting that contamination due to staggered adoption of mandated coverage is not a threat to our estimates.

[2] Buckles (2013) uses data from 1980-2002 and includes the states with mandates in the 2000s (Connecticut (2005) and New Jersey (2001)) in their control group. We use data from 1974-2014 and include states with mandates in the 2000s in our treatment group.

Table E.1: Effects of IVF coverage generosity level on multiple births per hundred live births, DD and DDD models

	Difference-in-Differences												Difference-in-Difference-in-Differences			
	All women				Women 35 and older				Women under 35							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
All levels	0.10*	0.07***			0.17	0.12			0.06	0.05***			0.49***	0.25***		
	(0.05)	(0.02)			(0.12)	(0.08)			(0.04)	(0.02)			(0.09)	(0.09)		
Level 0			0.01	0.01			-0.04	-0.03			0.01	0.01			0.45***	0.23**
			(0.06)	(0.03)			(0.15)	(0.10)			(0.04)	(0.03)			(0.12)	(0.11)
Level 1			-0.11***	-0.02			-0.30*	-0.30**			-0.10**	-0.01			0.22	-0.13
			(0.02)	(0.03)			(0.16)	(0.11)			(0.04)	(0.02)			(0.22)	(0.13)
Level 2			0.15***	0.17***			0.39***	0.31***			0.04***	0.11***			0.64***	0.40***
			(0.01)	(0.03)			(0.03)	(0.06)			(0.01)	(0.02)			(0.00)	(0.04)
Level 3			0.20***	0.06			0.37***	0.31***			0.14***	0.05			0.58***	0.39***
			(0.03)	(0.05)			(0.03)	(0.10)			(0.02)	(0.04)			(0.06)	(0.08)
Level 4			0.23**	0.20**			0.45***	0.33***			0.13**	0.16***			0.78***	0.54***
			(0.10)	(0.08)			(0.16)	(0.10)			(0.06)	(0.05)			(0.06)	(0.03)
Level 5			0.42***	0.19***			0.84***	0.62***			0.27***	0.12***			0.94***	0.73***
			(0.02)	(0.04)			(0.03)	(0.10)			(0.01)	(0.04)			(0.00)	(0.04)
Constant	1.00***	-3.23	0.99***	-1.75	1.34***	-1.45	1.33***	4.80	1.00***	-1.93	1.00***	-0.90	0.84***	7.14	0.84***	15.10***
	(0.02)	(2.22)	(0.01)	(2.51)	(0.06)	(7.54)	(0.06)	(7.60)	(0.01)	(2.06)	(0.01)	(2.43)	(0.04)	(5.58)	(0.04)	(3.42)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates included	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
P-value of $Ch2$ stat			0.00	0.00			0.00	0.00			0.00	0.00			0.00	0.00
Number of cells	1,768	1,768	1,768	1,768	1,768	1,768	1,768	1,768	1,768	1,768	1,768	1,768	3,616	3,276	3,616	3,276

Note: This table presents the DD, and DDD estimates from the effects of the generosity level of IVF coverage on multiple births per hundred live births. Data are aggregated into state-year cells for DD analysis and state-year-age cells for DDD analysis. All models include state and year fixed effects. Included covariates listed in notes for Table 4. Standard errors are clustered at the state level and appear in parentheses. The $Ch2$ statistic is used to test the null hypothesis that the estimated coefficients are all equal.

$*p < 0.10, **p < 0.05, ***p < 0.01$

Table E.2: Estimated effects from a DD model replicating Buckles (2013)

(a) Multiple births per hundred live births

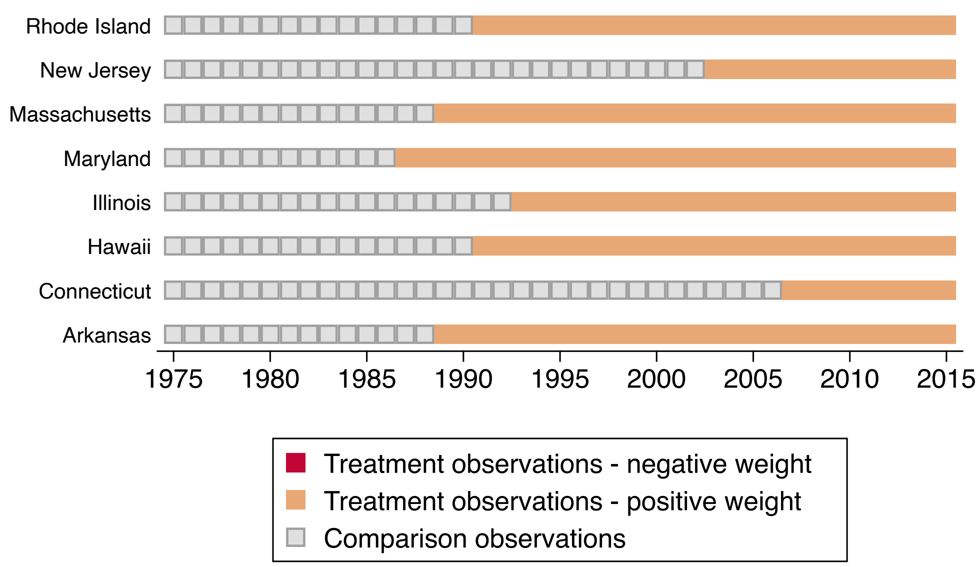
	All women		Women 35 and older		Women under 35	
	(1)	(2)	(3)	(4)	(5)	(6)
All levels	0.05	0.03	0.07	0.02	0.03	0.03*
	(0.06)	(0.02)	(0.12)	(0.07)	(0.03)	(0.01)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Covariates included	No	Yes	No	Yes	No	Yes
Observations	1,059	1,059	1,065	1,065	1,065	1,065

(b) Number of infants per thousand live birth

	All women		Women 35 and older		Women under 35	
	(1)	(2)	(3)	(4)	(5)	(6)
All levels	0.63	0.34	0.86	0.27	0.37	0.31*
	(0.64)	(0.22)	(1.27)	(0.78)	(0.39)	(0.15)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Covariates included	No	Yes	No	Yes	No	Yes
Observations	1,059	1,059	1,065	1,065	1,065	1,065

Notes: This table replicates estimates from Buckles (2013). The study sample includes birth certificate data from 1980-2002. The estimates compares the mandated states with never mandated states using the the DD model specified in E.1. Standard errors are clustered at the state level and appear in parentheses.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Figure E.1: Weights used in a Difference-in-Differences (DD) model by state and year fixed effects



Note: This figure characterizes the weights used in estimating a DD model with time and state fixed effects from the impact of mandated IVF coverage on multiple births per hundred live births and number of infants per thousand live births (identical weights for both outcome variables). The weights are the residuals from a regression of treatment on state and year fixed effects, scaled by the sum of the squared residuals across a pooled sample of mandated and never mandated states. See de Chaisemartin and D'Haultfoeuille (2020) for discussion.