Unintended Consequences of Policy Interventions: Evidence from Mandated Health Insurance Coverage for IVF Treatment*

Arezou Zaresani† Lucie Schmidt‡

November 2021

(Download the latest draft from http://azaresani.com/work-in-progress)

Abstract

Mandated health insurance coverage for expensive In Vitro Fertilization (IVF) varies widely in generosity across the US states. We find that more generous coverage within the states that mandate any coverage causes an increase in risky and costly multiple births. While more generosity is associated with fewer embryos transferred, that effect is dominated by changes in the composition of patients, where more older women with lower fertility pursue treatment. This is mirrored by a greater decline in child adoption to older women in states with more generous coverage. Compositional effects imply that increased access without regulation might impose additional burdens on the healthcare system.

JEL classification: G22; I11; I13; I18; J13; J16.

Keywords: healthcare costs; healthcare utilization; health insurance; mandated benefits; in vitro fertilization; child adoption.

*We thank Joelle Abramowitz, Pamela Campa, Jeff Clemens, Herb Emery, Mark Harris, Ian Heffernan, Lars Kirkeboen, Giulia La Mattina, Stephan Litsching, Arvind Magesan, Miguel Olivo-Villabrille, Juan Pantano, Elizabeth Savage, Hitoshi Shigeoka, Stefan Staubli, and Lindsay Tedds for their helpful comments. We have also benefited from discussion with conference and seminar participants at the Junior Health Economics Workshop (2020), University of Manitoba (2020), Graduate Research Institute for Policy Research (GRIPS 2020), Royal Melbourne Institute of Technology University (2019), American Economics Association Meeting (AEA 2019), Canadian Health Economics Study Group (CHESG 2018), Canadian Economics Association (CEA 2018), Australian Workshop on Econometrics and Health Economics (2018) and American Society of Health Economics (ASHEcon 2016).

†University of Manitoba, IZA and Tax and Transfer Policy Institute (TTPI) at Australian National University, Email: arezou.zaresani@umanitoba.ca. I acknowledge financial support from the University of Calgary, the University of Victoria, and University of Melbourne.

‡Smith College and NBER, Email lschmidt@smith.edu.

1 Introduction

Healthcare spending in the US has risen rapidly, from 5 percent of GDP in 1960 to 17.9 percent in 2017 (CMMS, 2017). Lifestyle changes and an aging population have contributed to increases in chronic illnesses such as cancer, musculoskeletal conditions, diabetes, and heart disease. These conditions have expensive treatment options, raising concerns about access to treatment and its overall costs. Policy interventions that mandate health insurance coverage for expensive medical treatments can affect patients' choices in several ways. More generous coverage can affect existing patients' utilization behavior by decreasing the cost of additional expensive treatments. However, more generous coverage could also expand access to new patients who might have used a cheaper alternative, further contributing to increases in healthcare costs. This could be important, particularly if more generous coverage leads to changes in the composition of patients seeking treatment such that patients with lower probabilities of success initiate treatment. Patients' behavioral responses to the increased accessibility of expensive treatments are critical to understanding health policy interventions' ramifications.

Mandated health insurance coverage for in Vitro Fertilization (IVF) treatment in the US provides an appealing case study for several reasons. First, the generosity of mandated coverage varies widely across states and over time. States range from no coverage to coverage of infertility treatments excluding IVF, to covering an unlimited number of IVF cycles, and mandates vary across several other dimensions of generosity as well.[1] This variation across states and time allows for the identification of effects of coverage generosity on patient's utilization and outcomes. Second, patients choose the intensity of their treatment (through the number of transferred embryos) based on their preferences and the expected costs and benefits. This choice directly affects both success rates and the likelihood of risky and costly multiple births. Finally, IVF resembles other medical treatments like those for heart disease or cancer, which are expensive and have uncertain outcomes (Shapiro and Recht, 2001).

[1] Other dimensions include age restrictions, coverage of unmarried women, coverage of cycles with frozen or donor eggs, minimum years of infertility to qualify for coverage, and whether mandates apply to Health Maintenance Organizations (HMOs). For more information see Table 1.

In this paper, we empirically investigate how the generosity of mandated coverage for IVF treatment affects patients' utilization behavior and the composition of those utilizing the treatment. In the absence of data on the utilization of IVF treatment for all years, we examine multiple births as a proxy for IVF treatment intensity.[2] More generous coverage could have two competing effects on the incidence of multiple births. First, existing patients face less pressure to conceive in each cycle, so they might choose less intense treatments and transfer fewer embryos (Jain et al., 2002; Reynolds et al., 2003). This could decrease the incidence of multiple births. Second, generous mandates could change the composition of the patients by expanding access to new patients who might not have pursued treatment in the absence of insurance coverage. This could lead to an increase in the incidence of multiple births.[3] The overall effect of more generous coverage for IVF treatment on the incidence of multiple births is, therefore, ambiguous. The increase in the incidence of multiple births from the latter effect could dominate the decrease from the former, especially if the new patients are mostly those with a lower probability of success who will transfer more embryos per cycle.

We use a Generalized Synthetic Control (GSC) model (Xu, 2017) to estimate the causal effects of the generosity of IVF coverage on the incidence of multiple births. This helps to account for both the possibility that the standard Difference-in-Differences (DD) parallel trends assumption might not hold and for problems associated with the treatment turning on at different times in different states.[4] We use administrative birth certificate data from the National Center for Health Statistics Detail Natality File on all births in the US between 1975 and 2014 and exploit variation in generosity levels of mandated coverage across states and over time. To shed light on patients' utilization behavior, we use fertility clinic data from the Society for Assisted Reproductive Technologies (SART) to examine the association between more generous IVF coverage and the composition

[2]See Section 3 for a discussion of the weaknesses of this measure and section 4.2 for regression discontinuity results that provide suggestive evidence that increases in multiple births are directly linked to coverage for IVF treatment.

[3]Bundorf et al. (2007) and Hamilton et al. (2018) refer to these two effects respectively as intensive and extensive margin effects, but not in the context of differing generosity levels within the set of states which mandate coverage.

[4]See Goodman-bacon (2019) for the exposition of the latter problem, and Ben-Michael et al. (2019) for a discussion of how synthetic control models can solve it.

of the pool of patients, the number of initiated IVF cycles, and the number of embryos transferred per cycle. We also use data from the National Data Archive on Child Abuse and Neglect (NDACAN) to examine the association between the generosity of IVF coverage and adoptions of children aged 0-6; as such adoptions could be considered in some circumstances a substitute for conceiving through IVF.

Three main findings emerge from our empirical analysis. First, after controlling for state-level characteristics, more generous IVF coverage causes an increase in the incidence of multiple births. For instance, in Massachusetts, which covers an unlimited number of cycles, multiple births per hundred live births increase by 27%, while in Arkansas and Hawaii, states that cover only one cycle, multiple births only increase by 8%. These effects are larger for older women: in Massachusetts, the increase is 44% for women over 35, compared with 21% for younger women. Second, states with more generous coverage have fewer average transferred embryos per cycle. This is true for both older and younger women. Finally, states with more generous coverage also see significant increases in the share of cycles performed on older women with lower fertility as new patients are drawn into treatment. This is mirrored by a greater decline in child adoption to older women in states with more generous IVF coverage. Our findings suggest that the change in the patients' composition dominates the reductions in embryos transferred per cycle, leading to the overall increases in the incidence of multiple births that we estimate.

Our paper is related to the literature investigating the effects of mandated coverage for infertility treatments on a variety of outcomes, including utilization of treatment, infant health outcomes, fertility, age at first birth, time of marriage, women's choice to pursue professional careers, and labor supply over the life cycle (Schmidt, 2005; Bitler and Schmidt, 2006; Bundorf et al., 2007; Schmidt, 2007; Bitler, 2007; Bitler and Schmidt, 2012; Abramowitz, 2014, 2017; Kroeger and La Mattina, 2017; Lundborg et al., 2017; Gershoni and Low, 2020a,b). Most of these studies use either state-year or state-year-age variation in mandated IVF coverage in Difference-in-Differences (DD) and Difference-in-Difference-in-Differences (DDD) frameworks, respectively.[5]

[5]There are two exceptions: Machado and Sanz-de-Galdeano (2015) uses a synthetic control model to estimate effects of mandated IVF coverage in the US on the timing of first births and women's total

Studies that relate most closely to our work examine the effects of the IVF mandates on the incidence of multiple births. Most of these studies find that mandates increase multiple births and are associated with worse health outcomes in terms of birth weight and gestation (Bundorf et al., 2007; Bitler, 2007; Kulkarni et al., 2013). Buckles (2013) finds that mandated coverage for IVF treatment has a small positive and statistically insignificant impact on the incidence of multiple births. Studies that use fertility clinic-level data find that treated patients with health insurance plans covering IVF treatment transfer fewer embryos than those with no insurance coverage (Jain et al., 2002; Reynolds et al., 2003; Henne and Bundorf, 2008; Hamilton and McManus, 2012). However, this previous literature ignores the differences in generosity within the set of states that mandate coverage for IVF.

Our main contribution to this literature is to provide the first comprehensive study on how patients' utilization responds to the *generosity* of mandated IVF coverage by combining various data sources. If policy makers wish to enact mandated coverage for IVF, one important element of policy design is the generosity of coverage, and the previous literature on the effects of the mandates does not differentiate between different levels of coverage. We further study the effects of the mandates on patients' utilization behaviors using the sample of all clinics, allowing us to explore the effects of the coverage's generosity where most of the previous work focuses on patients in one specific clinic within a state. We also provide the first evidence on the spillover effects of mandated IVF coverage on child adoption as a close alternative option to conceiving own infant. Our work also allows analysis of the more recent mandates legislated in the 2000s, which were not covered in much of the previous literature. Finally, our work benefits from methodological advances relative to the straightforward two-way fixed effects Difference-in-Differences methods used in previous work.

Our findings suggest that changes in the composition of the pool of patients are important in understanding the policy implications of increased health insurance generosity. This is consistent with previous studies on the role of incentives in healthcare utiliza-

fertility rates. Lundborg et al. (2017) uses IVF treatment as an instrumental variable to women's fertility decisions and examine the effects of having children on Danish women's careers.

tion. Chernew et al. (2000) suggest that patients should pay higher out-of-pocket costs for more expensive treatment in an optimal insurance plan. Einav et al. (2016) (in the case of breast cancer treatments) and Hamilton et al. (2018) (in the case of infertility treatments) both suggest that top-up pricing for more aggressive treatments could be optimal. Consistent with the work by Hamilton et al. (2018), Bhalotra et al. (2020) find that a Swedish single embryo transfer policy reduced the incidence of multiple births and improved maternal and infant health.

2 Background

2.1 IVF treatment

Infertility, defined as the inability to conceive or carry a pregnancy to full term, is recognized as a disease by both the American Society for Reproductive Medicine and the World Health Organization. Infertility treatment usually begins with medical tests and physician advice, often followed by the woman's use of one of several drugs to stimulate egg production. If these less expensive treatment methods are not successful, assisted reproductive technologies such as IVF treatment are often recommended. Success rates of a single IVF cycle are as low as 20 percent (CDC, 2015), and many patients require more than one cycle of treatment to achieve a live birth. One cycle of IVF treatment costs can be as high as 46 percent of the average US family's annual disposable income (Kissin et al., 2016).

In IVF treatment, eggs are extracted, a sperm sample is obtained, and eggs and sperm are manually combined. The fertilized eggs, called embryos, are then transferred into the woman's uterus.[6] The practice committee of the American Society of Reproductive Medicine provides guidelines on the maximum number of embryos to transfer per cycle (Klitzman, 2016).[7] However, given the high costs and low success rates of IVF, patients often wish to exceed these guidelines to improve their odds of success, and in doing so,

[6] The first infant conceived using an IVF treatment was born in 1978 in the UK.

[7] Currently, recommendations are for 1-2 embryos per cycle for women under the age of 35 and increase with age.

increase the likelihood of multiple births. Most monetary costs of multiple births are covered by insurance. Many patients with fertility problems view multiple births as a desirable outcome (Gleicher and Barad, 2009). However, multiple births are costly and risky for both mothers and infants (Merritt et al., 2014; Caserta et al., 2014).[8]

2.2 Mandated IVF coverage in health insurance plans

Due in part to concerns about the high cost of IVF treatment, between 1978 and 2005, 15 states in the US passed legislation pertaining to coverage of infertility treatment in employer-provided[9] private health insurance plans.[10] In these *mandate to cover* states, private health insurance companies are required to cover infertility treatment in all of their policies.[11]

The level of coverage in the mandate to cover states is quite heterogeneous. During our study period, Montana, New York, Ohio, and West Virginia mandate coverage for some types of less invasive infertility treatments but do not require coverage of IVF. Arkansas and Hawaii mandate coverage for only one cycle of IVF; Connecticut mandates up to two cycles; Rhode Island and Maryland mandate up to three cycles; Illinois and New Jersey mandate up to four cycles; and Massachusetts has no limit. Mandates also vary along with several other dimensions, including (but not limited to) age restrictions, coverage of unmarried women, minimum years of infertility to qualify for coverage, coverage for cycles with frozen or donor eggs, and whether mandates apply to health maintenance organizations (HMOs). However, as shown in Table 1, these dimensions of generosity are highly correlated with the mandated number of cycles, so we treat the number of cycles as a proxy for the overall generosity level of mandated coverage. There are 35 states

[8] The average cost of a singleton birth was $27,000 in 2012, while twin and triplet births cost $115,000 and $435,000, respectively (Lemos et al., 2013). The risks of multiple births to mothers include high blood pressure, gestational diabetes, and a higher cesarean section rate. The risks to infants include low birth weight, prematurity, and sometimes long-term disabilities like autism and cerebral palsy (Hoffman and Reindollar, 2002; Fritz, 2002; Martin and Park, 1999; Reynolds et al., 2003).

[9] Under the 1974 Employer Retirement Income Security Act (ERISA), self-insured firms are exempt from these mandates.

[10] We extract the mandated coverage date and the details of the coverage from the National Infertility Association website. For more information see `https://resolve.org`.

[11] In *mandate to offer* states, health insurance companies are required to offer plans that would cover infertility treatment but are not required to include this coverage in all policies. We exclude these states (California, Texas, and Louisiana) from our empirical analysis.

which never legislated policies to mandate coverage for infertility treatments.[12] These *never mandate states* serve as a control group in our analysis.

3 Data

We use several data sources for our empirical analysis. First, we use administrative birth certificate data from the National Center for Health Statistics Detail Natality Files. The data comprise records of live births in the US from 1975 to 2014, and include parental information such as mother's age, education, and race, father's race, parental marital status, and state of residence; and infant information such as sex, birth order, and the plurality (single or multiple births). Our study sample includes the 12 mandate to cover states (treatment group) and the 35 never mandate states (control group). We aggregate the data into state-year cells for our empirical analysis.[13]

Our primary outcome variable is the multiple birth rate, defined as the number of multiple births (i.e., not singletons) per hundred live births.[14] Multiple births are a useful proxy for the aggressiveness of treatment. More than one-third of twins and more than three-quarters of triplets and higher-order multiples in the US in 2011 resulted from conceptions assisted by infertility treatments (Kulkarni et al., 2013). However, one caveat of this approach is that in the birth certificate data, we have no way of knowing whether the multiple births are naturally occurring, due to IVF treatment, or due to other infertility treatments besides IVF.[15] The simple multiple birth indicator also does not differentiate between a twin birth and a quadruplet birth, even though these have

[12] Since the end of our study period in 2014, 4 additional states have mandated IVF coverage: Colorado (2020), New Hampshire (2020), New York (2020), and Delaware (2018). We do not include these mandates in our analysis but include these states in our control group.

[13] The public-use birth certificate data include the mother's state of residence only through 2004, so we use restricted access data files from 2005 to 2014. A few states do not report some parental information in some years. We impute these missing values in the state-year aggregated data by setting them to the corresponding variable's average in the years before and after.

[14] There is one record for each infant in the data file (e.g., there are three records for a triplet birth). The number of infants, therefore, over-represents the incidence of multiple births. To deal with this issue, we follow Buckles (2013) and construct a weight by dividing one by the plurality of each infant (i.e., the weight of each infant in a triplet birth is set as 1/3). We use these weights to convert the unit of analysis from infant to birth.

[15] The birth certificate data includes a variable indicating births with assisted reproductive technology starting from 2011. However, the variable has many missing values and is not very informative.

very different cost implications, so we also examine the effects of generosity on the number of infants per thousand live births.

Second, we use the March Annual Social and Economic Supplement of the Current Population Survey (CPS) to create control variables at the state-year level, including the population percentage of women of childbearing age, the female labor force participation rate, and real per capita income.[16] To account for the share of women who will be affected by the mandates, we control for the percentage of working-age individuals with private health insurance, as well as the percentage of working-age individuals in large firms (defined as those with +500 employees) as a proxy for the share of workers in self-insured firms and therefore not subject to the mandates under the Employer Retirement Income Security (ERISA) act.[17]

Third, we use fertility clinic-level data collected from 1996 to 2010 by the Society for Assisted Reproductive Technology (SART) to study patients' utilization of IVF treatment.[18] The data include information on the number of cycles initiated in each clinic, the share of cycles performed on women 35 and older, and the average number of embryos transferred by mothers' age. We exclude frozen, and donor cycles since only fresh and non-donor cycles are covered by mandates in many states.

Finally, we use data on child adoptions from the National Data Archive on Child Abuse and Neglect (NDACAN) from 2000 to 2014.[19] The data includes the records of all the public adoptions in the US and has information on adoptive parents' age and race and the adopted children's age, sex and race, and the year and the state the adoption is finalized. We focus on 0-6 years old children since younger children might be closer substitutes for newborn infants. We aggregate the data into state-year cells and create a variable representing the number of young adopted children per one thousand live births

[16]We convert all dollar values to 2007 dollars using the Consumer Price Index.

[17]Large firms are more likely to self-insure (Gabel et al., 2003; Park, 2000).

[18]SART has a voluntary reporting system and about 10% of the clinics do not report data. SART does not regulate clinic practices. The available range of data does not cover the 1980s and earlier 1990s IVF mandates.

[19]The data is collected under a federally mandated system for all children in foster care and on children adopted under the auspices of the state public child welfare agency. The available range of data does not cover the 1980s and earlier 1990s IVF mandates.

in that state and year.[20]

4 Empirical analysis

4.1 Descriptive evidence

Table 2 presents summary statistics from the birth certificate data from 1975 to 2014, presented in ten-year intervals and broken out by IVF mandate status. In more recent years, mothers are, on average, older, more educated, and less likely to be married. Multiple births per hundred live births (multiple birth rate) and the number of infants per thousand live births are also higher in recent years. The incidence of multiple births in states with mandated coverage is higher than that in the never mandate states, and this gap is widening over time.

The age of 35 is considered a turning point in women's fertility: one-third of women older than 35 experience fertility problems (CDC, 2015). Therefore, we present all of our empirical analyses first for all women, then separately by women 35 and older and women younger than 35 years. Figure 1 plots trends in multiple births per hundred live births by generosity level of mandated IVF coverage, first for all women, then separately for older and younger women. Three main patterns emerge. First, the incidence of multiple births is increasing across all states over our study period. Second, more generous coverage is associated with more rapid growth in the incidence of multiple births. Third, the association between coverage generosity and the incidence of multiple births is stronger for older women than for younger women.[21]

4.2 Is the increase in multiple births driven by IVF treatment?

Figure 2 shows that the multiple birth rate increases with women's age, and this pattern is stronger in recent decades. Older women are more likely to have multiple births, even

[20]Our data do not include private adoptions (either domestic or international). Our analyses of the insurance mandates' effects will be biased if the generosity of mandated IVF coverage differentially affects private adoptions versus those through the state welfare system.

[21]The patterns for the number of infants per thousand live births are similar.

in the absence of infertility treatment, but the increase in the incidence of multiple births in recent decades reflects an increase in infertility treatments. To examine the extent to which IVF coverage might be responsible, we compare the multiple birth rates of women eligible for mandated IVF coverage with the rates for ineligible women within the same state.

Women over 40 years old in Connecticut and Rhode Island and women over 46 years old in New Jersey are not eligible for mandated coverage (see Table 1).[22] We explore the sharp discontinuity in the eligibility ages for the mandated coverage in these three states using a Regression Discontinuity Design (RDD) model, using women's age as the running variable. We compare the incidence of multiple births to women right above the age eligibility threshold who are not eligible for mandated coverage to those right below the threshold who are eligible. We estimate a regression of the form:

$$y_{ia} = \delta + f(a) + \rho D_a + \beta X_i + \epsilon_i \tag{1}$$

where y_{ia} denotes whether the birth to woman i with age a is a multiple birth. D_a is the treatment dummy that switches on for women above the eligibility age threshold. X_i is a set of individual characteristics, including women's race, education, and marital status. As noted above, older women are more likely to have multiple births even in the absence of treatment, so $f(a)$ denotes a linear age trend to control for this age variation. ϵ_i is the error term. The coefficient of interest is ρ, which captures the intent-to-treat effect of the mandated coverage on the multiple birth rate. The identification assumption is that the other unobservable variables affecting the incidence of multiple births change smoothly in the neighborhood of the age eligibility threshold (Hahn et al., 2001).

We follow Schmidt (2007) and allow mandated coverage to affect multiple births with a two-year delay. This delayed effective mandated coverage year accounts for two factors: first, infertility treatments may not lead immediately to conception, and second, a successful conception will not translate into a birth until nine months later. Therefore, we use the birth certificate data from two years after the mandated coverage in each

[22] Only these there states have age limits for the mandated IVF coverage.

state up to 2014 and focus on women within the 5-years window around the eligibility age for estimating Equation (1).[23] The bandwidth and degree of the fitted polynomial are selected using the method of Calonico et al. (2020), and the standard errors are clustered at the age level. Table 3 shows that after controlling for individual characteristics and a linear trend in age, the incidence of multiple births for women not eligible for the mandated coverage decreases by 7.29%, 5.94%, and 5.55% respectively in Connecticut, Rhode Island, and New Jersey. We then estimate the effects on the incidence of multiple births in New Jersey from a placebo eligibility age threshold at 40 years.[24] The last panel of Table 3 shows that the estimated effect is negligible and insignificant. Overall, the findings from our RDD analysis provide evidence that mandated coverage for IVF treatment is a driving factor in the increase in the incidence of multiple births in the states with mandated coverage.

4.3 Mandated IVF coverage and incidence of multiple births

We use a Generalized Synthetic Control (GSC) framework developed by Xu (2017). Much of the previous literature on infertility mandates use upon Difference-in-Difference models, which rely on the assumption that the trends between treatment and control groups would have been parallel in the absence of the policy change. While this might be true for the adoption of any mandate, Figure 1 suggests that it might be violated when looking specifically at the mandates' generosity. In addition, recent methodological advances illustrate problems with the standard Difference-in-Differences or two-way fixed effects models when policy variation turns on in different states at different times (Goodman-bacon, 2019). Our GSC framework helps with both issues (see Ben-Michael et al. (2019) for a discussion of the use of GSC models in this context).

The GSC model is a generalization of the conventional synthetic control models using

[23] The sample includes all births to women ages 35 to 45 in Connecticut between 2007 (two years after mandated IVF coverage in 2005 for women below 40 years) and 2014, and in Rhode Island between 1991 (two years after mandated IVF coverage in 1989 for women below 40 years) and 2014, and all births to women ages 41 to 51 in New Jersey between 2003 (two years after mandated IVF coverage in 2001 for women below 46 years) from the birth certificate data.

[24] The placebo estimates in New Jersey uses data on all the births to 35 to 45 years old women (an age window with no change in eligibility for IVF coverage) between 2003 (two years after mandated IVF coverage for women below 46 years old) and 2014.

a linear interactive fixed-effect framework, in the spirit of the weighting scheme of the original synthetic control method developed by Abadie et al. (2010).[25] A GSC model uses the control group and the treatment group (in pre-treatment periods) to impute treated counterfactuals. We estimate a model of the form:

$$y_{it} = \delta_{it} D_{it} + \beta X'_{it} + \lambda'_i f_t + \epsilon_{it}, \tag{2}$$

where i and t respectively denote state and time and y_{it} denotes the outcome variable. Our main outcome variables are the multiple births per hundred live births and the number of infants per thousand live births. D_{it} is a dummy variable coded as one for treated state i in years following the mandated coverage.[26]The vector X_{it} is a set of time-varying state-level characteristics, including mothers' age, marital status, education, mothers' and fathers' race. We also include the state-level socioeconomic characteristics from the CPS data discussed earlier.[27]

$\lambda'_i f_t$ denotes the interactive fixed effects where λ_i and f_t are r-vectors of respectively state-specific intercepts and time-varying coefficients, capturing unobserved common factors that can be decomposed into a state-year multiplicative form. This interactive component covers a wide range of unobserved heterogeneity, but it does not capture unobserved confounders that are independent across states. ϵ_{it} is the error term and captures any remaining unobserved components that affect the outcome variable. r is estimated through a data-driven procedure where a larger value covers a broader range of unobserved heterogeneity. Intuitively, a GSC framework allows the data to tell which model fits better.[28] Details of the estimation procedure of a GSC framework are provided in

[25]There are two main approaches to estimate causal effects when the common trend assumption is likely to be violated. The first approach uses a matching method to condition on pre-treatment observable characteristics (Abadie, 2005; Abadie et al., 2010, 2015). This approach helps to balance the effects of time-varying confounders between the treatment and control groups. The second approach explicitly models the unobserved time-varying confounders using an interactive fixed-effect model, which includes state-specific intercepts interacted with time-varying coefficients (?). GSC links the matching and interactive fixed-effect methods and brings together synthetic control and interactive fixed-effect models where the Difference-in-Differences model is a special case. For a review of recent studies on synthetic control methods, see Abadie (2020).

[26]We follow Schmidt (2007) and allow the mandated coverage to affect the incidence of multiple births with a two-year delay. See Section 4.2 for the reasons.

[27]For more information, see Section 3.

[28]For instance, for $r = 2$ if we set $\lambda'_i = (1, \alpha_i)$ and $f'_t = (\tau_t, 1)$ then $\lambda'_i f_t = \alpha_i + \tau_t$. In this case, the

Appendix B.

The coefficients of interest are δ_{it} which capture the treatment effect on treated state i at time t. The average treatment effect on treated is the average of all the treated states' estimates. We use data from a 15-year window around the effective mandated coverage year (15 pre- and 15 post-treatment periods) for our estimations.[29] We aggregate the data into state-year cells and estimate the model separately for each generosity level indicated in Table 1. Standard errors are estimated using a parametric bootstrapping procedure using 2,000 re-sampling draws of the residuals (Xu, 2017).

The GSC framework has several advantages relative to the original synthetic control method by Abadie et al. (2010). First, it allows for more than one treated state with variable treatment periods. Second, it provides estimates of standard errors and confidence intervals, making inference more reliable. Third, it provides a data-driven procedure to select the number of factors in an interacted fixed-effect model (r) to minimize the perdition error and reduce over-fitting risk. Furthermore, this approach enables us to take advantage of the long pre-treatment panel to decrease the bias of the estimated effects.[30]

To pool states with similar generosity levels of coverage but different enacted mandate dates (i.e., level 1 coverage states: Arkansas (mandate date: 1987) and Hawaii (1989)), we assume that the responsiveness to the mandated coverage at the relative time to the mandated coverage is similar across the states, such that our analysis picks up the differences in generosity levels of the mandated coverage. This assumption is plausible since we control for flexible state and time fixed effects and time-varying state-level characteristics affecting IVF treatment's utilization and outcome.

4.4 Results

Plots presenting the estimated counterfactual and estimated effects on the treated states for each level of coverage are presented in Figure 3 and suggest that the GSC estimator

GSC model is reduced to a conventional DD model with state and time fixed effects.

[29]Exceptions are Connecticut (mandate enacted in 2005) with an 8-year post-treatment period and Hawaii (1987) and Arkansas (1987) with 10-year pre-treatment periods because the 15-year window for these states falls outside our data availability period of 1975–2014.

[30]See Abadie (2020) for a review of recent synthetic control methods.

works quite well in imputing counterfactuals for the treated states to match the control group in the pre-treatment periods. Table 4 presents the estimated effects of the generosity level of mandated coverage on multiple births per hundred live births. The first set of columns presents the estimated effects for all women. Panel A presents the estimates using one indicator that pools all mandate to cover states, regardless of generosity level, to be consistent with the previous literature on infertility mandates. The first column shows that any mandated coverage increases the multiple birth rate by 0.10 percentage points relative to the never mandated states, an 8.84% increase from a mean value of 1.13. The second column adds covariates to the model, which reduces the magnitude of the estimated effect to a 0.05 percentage point (4.42%) increase in the multiple birth rate.

Panels B through G show estimated effects broken out by level of generosity. Panel B shows that coverage for less invasive infertility treatment only (level 0) does not affect the multiple birth rate relative to states that never enact mandates. This finding is relatively consistent across our results. Panel C through G show that, in general, states with more generous coverage exhibit larger increases in multiple birth rates. Estimated effects with covariates range from a 0.08 percentage point increase (8.33%) in states with level 1 coverage to a 0.28 percentage point increase (26.92%) in states with level 5 coverage.

The remaining columns of Table 4 present the estimates for women 35 and older versus younger than 35, and the GSC plots by women's age are presented in Figure 4 and Figure 5. After controlling for covariates, the estimated effects for women 35 and older tend to be larger than those of younger women, especially at higher coverage levels. For older women, the estimated effects after controlling for covariates vary from no significant effect in states with level 1 coverage to a 0.56 percentage point (44.09%) increase in states with level 5 coverage. The estimated effects for younger women, especially for high generosity states, are much smaller (for example, a 0.21 percentage point (20.59%) increase in level 5 states).[31]

While our multiple birth rate measure indicates whether the birth included more than

[31] We also estimated the effects separately for each state with mandated coverage. These effects (available upon request) are quite similar to those aggregated by the coverage level.

one infant, our alternative outcome measure, the number of infants per thousand births, allows, for example, triplets to count more than twins. Table 5 presents the effects of the generosity level of mandated coverage on the number of infants per thousand live births. The overall findings are pretty consistent with those from the multiple births rate specification. After controlling for covariates, the estimated effect of any mandated coverage (Panel A) is 0.64 additional infants per thousand live births (a 5.51% increase). The estimated effects by the generosity level of coverage after including covariates range from 0.91 infants (9.37%) in states with level 1 coverage to 2.92 infants (27.68%) in states with level 5 coverage, and again, the effects are larger for older women.

Overall, our estimates from the GSC framework show that mandated coverage causes an increase in the incidence of multiple births, that states with more generous coverage experience larger estimated effects, and that effects are larger for women over 35 years.[32]

4.5 Robustness analysis

We estimate the effects of the generosity level of mandated coverage on the incidence of multiple births using a DD framework by exploring variations across states and over time. In addition to robustness check of our findings from the GSC framework and facilitating comparison with the previous literature, this analysis updates findings of Buckles (2013) –which uses data from 1980 to 2002– using more recent data from 1975 to 2014, allowing us to include two states with more recent mandated coverage; Connecticut and New Jersey, which mandated coverage in 2005 (2 cycles) and 2001 (3 cycles), respectively.

To further investigate the robustness of our findings from the GSC framework, we estimate the effects of the mandated coverage on the incidence of multiple births using a DDD framework. In our DDD analysis, we further refine the treatment group by mothers'

[32]There are other dimensions besides age that are strongly associated with infertility and IVF utilization, including education, marital status, and race (Bitler and Schmidt, 2006). College-educated women face incentives to postpone childbearing and invest in their professional careers and are more likely to work in jobs that offer private health insurance. Married women struggling with fertility seek infertility treatment, especially IVF, more often than unmarried women, and some mandated coverage explicitly excludes unmarried women. Although white women are less likely to experience infertility than black women, they are more likely to seek infertility treatment. We estimated the effects of the generosity of mandated coverage on the incidence of multiple births along these dimensions, and the results are mostly consistent with the patterns found in the previous literature. These estimates are available from authors upon request.

age. We use mandated coverage variations over the state, year, and women's ages (below and above 35 years old). This analysis allows us to control for two kinds of potentially confounding trends. First, we control for any time trends in the incidence of multiple births for women of a particular age that are constant across states. Second, we control for differences across states in the incidence of multiple births that affect all mothers, possibly due to other state policies or state-level economic conditions that might affect women's fertility decisions.

We aggregate the birth data into state-year and state-year-age cells for estimating the DD and DDD models, respectively. Specifications of the models and the estimated effects on multiple births per hundred live births and the number of infants per thousand live births are presented in Appendix C.

Our estimates from the DD model are statistically significant and larger in magnitude than the estimates of Buckles (2013).[33] This finding could be driven by the states with the most recent and more generous mandated coverage, which were not included in Buckles (2013). The overall story is the same as our findings from the GSC framework; more generous coverage is associated with an increase in the incidence of multiple births, and this association is stronger for older women.

5 Patients' behaviour

Our estimates from the GSC models show that more generous coverage causes an increase in the incidence of multiple births. This is despite speculation that more generous coverage might reduce the incidence of multiple births by reducing patients' incentives for transferring more embryos per cycle. However, more generous IVF coverage could also provide incentives to new patients with lower probabilities of success to seek treatment.

We use two additional data sources to shed light on patients' behavior from mandated coverage generosity. First, we investigate patients' utilization behavior using fertility

[33] Our estimated effects using a DD model 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 estimates of Buckles (2013) respectively at 0.02 (p-value > 0.10) and 0.28 (p-value > 0.10).

clinic-level data.[34] Second, we investigate child adoption as the main alternative to live birth. However, since data collection for both of these datasets started after several mandates were passed, these analyses should be thought of as descriptive and not provide causal estimates.

5.1 Evidence from IVF clinics

We use SART's clinic-level data from 1996 to 2010 to investigate how mandated coverage's generosity affects patients' utilization behavior. Table A.1 presents summary statistics. The average number of transferred embryos decreases over our study period in both mandated and never mandated states, likely due partly to changes in SART recommendations over time.[35] More embryos are transferred per cycle for women 35 and older than for younger women. The share of cycles performed on women 35 and older is ten percentage points higher in recent years in the mandated states relative to the never mandated states.

We cannot use our GSC model to examine how coverage's generosity affects patients' utilization behavior because the mandated coverage date for 6 out of 8 states falls before SART data is available. Including clinic or state fixed effects would absorb all the variation. We instead use a linear Mixed Effect (ME) model to estimate the relationship between coverage generosity and patients' utilization behavior. This approach takes advantage of the hierarchical structure of our data, with clinics nested within states.[36] We exploit random variation between clinics within states in addition to the variation across the states. An example of random variation between the clinics might be doctors' opinions about the appropriate number of embryos to transfer, which would affect the

[34] Most of the previous studies use patient-level data from a specific fertility clinic within a state and compare patients with insurance coverage for IVF with those with no coverage. Our data includes all clinics across all the states, allowing us to study the effects of the generosity of mandates on patients' utilization behaviors.

[35] A major change to SART's guidelines occurred in 2004. We estimated an event study model and find that this change is associated with reducing the number of embryos transferred for both younger and older patients. However, the estimated effects do not vary by coverage generosity. Details are provided in Appendix D.

[36] ME models are extensively used in education research where the independence assumption for causal inference in a linear model is violated; for instance, in studies where students and teachers are nested in classrooms, schools and districts (Goldstein, 1999).

incidence of multiple births from an IVF cycle. We estimate a model specified as:

$$y_{ist} = \alpha + \rho Level_{st} + \beta X'_{st} + \lambda_t + \nu\gamma_i + \omega\gamma_s + \epsilon_{ist} \tag{3}$$

where i, s and t denote clinic, state, and year, respectively, and y_{ist} denotes the outcome variable. Our outcome variables are the total number of cycles, the share of cycles performed on women 35 and older, and the average number of transferred embryos per cycle. $Level_{st}$ is an indicator for the generosity level of mandated coverage in state s at year t, with never mandated states as the reference group. The vector X_{st} includes the time-varying state characteristics from the CPS data described in Section 3. λ_t denotes year fixed effects, which pick up any factors changing over time that are common across the states and clinics (e.g., advances in IVF technology at the national level). γ_i and γ_s denote clinic and state random effects, respectively. ϵ_{ist} captures any remaining unobserved factors affecting the outcome variable. The coefficient of interest is ρ, which captures the relationship between the generosity level of mandated coverage and the outcome variable. ME models assume that first, clinic, and state-level residuals are uncorrelated; second, the errors –as measured by the residuals– at the state level are uncorrelated.

Table 6 presents the estimated effects for all women as well as results broken out by age. These results suggest the following: First, more generous coverage is associated with a significant increase in the average number of cycles in a clinic. Second, the relationship between the generosity of mandate coverage and the average number of transferred embryos per cycle shows that more generous coverage is associated with fewer transferred embryos for both older and younger women, with stronger effects for younger women. This would imply a decrease in the incidence of multiple births. Third, more generous coverage is associated with a higher share of cycles initiated by older women, which suggests changes in the composition of the patients seeking treatment. Given that older women transfer more embryos per cycle, this would imply an increase in the incidence of multiple births. Our GSC results using birth data show an overall causal increase in the incidence of multiple births, which suggests that the compositional effect may dominate.

5.2 Evidence from child adoption

Women who cannot naturally conceive an infant have two alternative pathways to motherhood: using IVF treatment or adopting a child. There is a significant overlap between these two options. More than half of the individuals who received infertility treatment had also considered adoption (Chandra et al., 2005). Gumus and Lee (2012) show that one-third of individuals who consider adoption have also sought IVF treatment. Both of these options have pros and cons. Despite technological advances, IVF treatment is expensive and has a low probability of success. Adopting a child is also expensive, uncertain, and can take a long time. Furthermore, some individuals might prefer to have a biological child. If more generous mandated coverage for IVF induces more older women to initiate IVF, we would expect that effect to be accompanied by a decrease in child adoptions.

Previous studies have examined the relationship between IVF treatment and child adoption. Gumus and Lee (2012) find that higher adoption rates at the state-year level are associated with a lower number of IVF cycles performed. Cohen and Chen (2010) find that mandated IVF coverage did not affect child adoption in mandated states relative to never mandated states. However, the effects of mandated coverage on adoption could be pretty heterogeneous, depending on the generosity of coverage and the women's age.

We use NDACAN's child adoption data from 1995 to 2014 to investigate the relationship between the generosity level of mandated IVF coverage and child adoption. We focus on 0-6 years old adopted children as such adoptions could be considered in some circumstances a substitute for conceiving through IVF. Table A.2 presents descriptive statistics. In the early years of our study period, the mandated states' adoption rate is higher than in the never mandated states. However, by the later half of our time period, this pattern had reversed itself so that the never-mandated states saw two more adopted children per ten thousand live births than did the mandated states.

Similar to our SART data analysis, we cannot use our GSC model because the mandated coverage date for 6 out of 8 mandated states falls before the availability of the adoption data. To examine the relationship between IVF coverage generosity and child adoption, we estimate an ME model similar to Equation (3), including time fixed effects and state random effects. The outcome variable is the number of 0-6 years old adopted children per ten thousand live births in each state-year cell. Table 7 presents the estimated effects, first for all women and then broken out by the age of adoptive women. Our results suggest a negative association between the generosity level of mandated coverage and the number of adopted children per ten thousand newborn infants that is much stronger for older women than for their younger counterparts.

Our analyses of these three different data sources, birth certificates, fertility clinics, and child adoption, have three main takeaways. First, more generous IVF coverage increases the incidence of multiple births. Second, more generous coverage is associated with a decrease in the number of transferred embryos for all women, but the association is stronger for younger women than older women. Third, more generous coverage is associated with changes in the composition of patients, where the share of cycles performed on women over 35 increases with coverage generosity. This is mirrored by a decrease in child adoption to older women in states with more generous coverage. These findings suggest that the increase in the incidence of multiple births from the change in the composition of patients is more substantial than the decrease in the incidence of multiple births from transferring fewer embryos per cycle, resulting in an overall increase in risky and costly multiple births.

6 Conclusion and policy implications

How do increases in the accessibility of expensive medical treatments affect patients' utilization behavior, and what are the resulting implications for healthcare costs? We explore the generosity of state-level mandated coverage for IVF treatment in the US. More generous coverage has been proposed as a way to decrease the incidence of risky multiple births by encouraging patients to transfer fewer embryos per cycle. While we find that more generous coverage is associated with transferring fewer embryos per cycle for all women, it also increases the share of cycles performed on older women. Overall, more generous coverage significantly increases the incidence of risky and costly multiple births. Our analysis highlights the importance of unintended consequences of the increased accessibility of an expensive medical treatment through changes in the composition of patients seeking treatment.

Our results are consistent with work by Bitler and Carpenter (2016), who show that mandated insurance coverage for mammography significantly increased mammography screenings and subsequently increased the detection of pre-cancers. However, they also find that a large share of the increased screenings resulted from utilization that was inconsistent with the American Cancer Society's recent recommendations. Our findings are also related to suggestions by Hamilton et al. (2018) –in the context of IVF– and Einav et al. (2016) –in the context of breast cancer treatment– for either regulating or limiting the aggressiveness of treatments; or for imposing a top-up price for more aggressive treatments; or some combination of the two. In the IVF context, Hamilton et al. (2018) argue that a value-based policy in which insurance plans cover single embryo cycles but patients must pay a top-up cost for transferring additional embryos could maximize welfare. This is consistent with findings of Bhalotra et al. (2020) from a Swedish single embryo transfer policy, which reduced the incidence of multiple births and improved maternal and infant health. Ignoring compositional effects could mean that increased access without regulation might impose additional burdens on the healthcare system.

References

Abadie, A. (2005). Semiparametric Difference in Differences Estimators. *The Review of Economic Studies*, 72(1):1–19.

Abadie, A. (2020). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*.

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490):493–505.

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

Abramowitz, J. (2014). Turning back the ticking clock: The effect of increased affordability of assisted reproductive technology on women's marriage timing. *Journal of Population Economics*, 27(2):603–633.

Abramowitz, J. (2017). Assisted Reproductive Technology and Women's Timing of Marriage and Childbearing. *Journal of Family and Economic Issues*, 38(1):100–117.

Ben-Michael, E., Feller, A., and Rothstein, J. (2019). Synthetic controls and weighted event studies with staggered adoption. Technical report.

Bhalotra, S., Clarke, D., and Mühlrad, H. (2020). Health and Labor Market Impacts of IVF Technology: Evidence from a Swedish Policy Mandate. Technical report.

Bitler, M. P. (2007). Effects of Increased Access to Infertility Treatment on Infant and Child Health : Evidence from Health Insurance Mandates. Technical Report February.

Bitler, M. P. and Carpenter, C. S. (2016). Health insurance mandates, mammography, and breast cancer diagnoses. *American Economic Journal: Economic Policy*, 8(3):39–68.

Bitler, M. P. and Schmidt, L. (2006). Health disparities and infertility: Impacts of state-level insurance mandates. *Fertility and Sterility*, 85(4):858–865.

Bitler, M. P. and Schmidt, L. (2012). Utilization of infertility treatments: The effects of insurance mandates. *Demography*, 49(1):125–49.

Buckles, K. (2013). Infertility insurance mandates and multiple births. *Health Economics*, 789(22):775–789.

Bundorf, M. K., Henne, M., and Baker, L. (2007). Mandated health insurance benefits and the utilization and outcomes of infertility treatments. Technical report.

Calonico, S., Cattaaneo, M. D., and Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *Econometrics Journal*, 23:192–210.

Caserta, D., Bordi, G., Stegagno, M., Filippini, F., Podagrosi, M., Roselli, D., and Moscarini, M. (2014). Maternal and perinatal outcomes in spontaneous versus assisted conception twin pregnancies. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 174(1):64–69.

CDC (2015). Centers for Disease Control (CDC) 2015 Fertility Clinic Success Rates Report.

Chandra, A., Martinez, G. M., Mosher, W. D., Abma, J. C., and Jones, J. (2005). Fertility, family planning, and reproductive health of U.S. women: Data from the 2002 National Survey of Family Growth. *Vital and health statistics*, 13(25):1–160.

Chernew, M. E., Encinosa, W. E., and Hirth, R. A. (2000). Optimal health insurance: The case of observable, severe illness. *Journal of Health Economics*, 19(5):585–609.

CMMS (2017). National Health Expenditure Summary Including Share of GDP, CY 1960-2017, Centers for Medicare and Medicaid Services. https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nationalhealthaccountshistorical.html.

Cohen, I. and Chen, D. (2010). Trading-off reproductive technology and adoption: Does subsidizing IVF decrease adoption rates and should it matter? *Minnesota Law Review*, 95:486–574.

Einav, L., Finkelstein, A., and Williams, H. (2016). Paying on the margin for medical care: Evidence from breast cancer treatments. *American Economic Journal: Economic Policy*, 8(1):52–79.

Fritz, M. A. (2002). Infertility treatment and the multifetal gestation epidemic: Too much of a good thing. *Contemporary OB/GYN*, 47(9):65–90,92.

Gabel, J. R., Jensen, G. A., and Hawkins, S. (2003). Self-insurance in times of growing and retreating managed care. *Health Affairs*, 22(2):202–210.

Gershoni, N. and Low, C. (2020a). Older yet fairer: How extended reproductive time horizons reshaped marriage patterns in Israel.

Gershoni, N. and Low, C. (2020b). The power of time : The impact of free IVF on women's human capital investments. Technical report.

Gleicher, N. and Barad, D. (2009). Twin pregnancy, contrary to consensus, is a desirable outcome in infertility. *Fertility and Sterility*, 91(6):2426–2431.

Goldstein, H. (1999). Multilevel statistical models. *London: Institute of Education.*

Goodman-bacon, A. (2019). Difference-In-Differences with Variation in Treatment Timing.

Gumus, G. and Lee, J. (2012). Alternative paths to parenthood: IVF or child adoption? *Economic Inquiry*, 50(3):802–820.

Hahn, J., Todd, P., Van der Klaauw, W., and ser Klaauw, W. V. (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, 69(1):201–209.

Hamilton, B. H., Jungheim, E., Mcmanus, B., Pantano, J., and Jungheim, E. (2018). Healthcare Access , Costs , and Treatment Dynamics : Evidence from In Vitro Fertilization. *American Economic Review*, 108(12):3725–3777.

Hamilton, B. H. and McManus, B. (2012). The effects of insurance mandates on choices and outcomes in infertility treatment markets. *Health Economics*, 21(8):994–1016.

Henne, M. B. and Bundorf, M. K. (2008). Insurance mandates and trends in infertility treatments. *Fertility and Sterility*, 89(1):66–73.

Hoffman, D. I. and Reindollar, R. H. (2002). Should we consider twins a success in OI and ART? *Controversies in OB/GYN*, (September).

Jain, T., Harlow, B. L., and Hornstein, M. D. (2002). Insurance Coverage and in vitro fertilization. *New England Journal of Medicine*, 332(9).

Kissin, D. M., Boulet, S. L., and Jamieson, D. J. (2016). Fertility treatments in the United States. *Obstetrics & Gynecology*, 128(2):387–390.

Klitzman, R. (2016). Deciding how many embryos to transfer: Ongoing challenges and dilemmas. *Reproductive Biomedicine and Society Online*, 3:1–15.

Kroeger, S. and La Mattina, G. (2017). Assisted reproductive technology and women's choice to pursue professional careers. *Journal of Population Economics*, 30(3):723–769.

Kulkarni, A. D., Jamieson, D. J., Jones, H. W., Kissin, D. M., Gallo, M. F., Macaluso, M., and Adashi, E. Y. (2013). Fertility treatments and multiple births in the United States. *New England Journal of Medicine*, 369(23):2218–2225.

Lee, M. S., Evans, B. T., Stern, A. D., and Hornstein, M. D. (2016). Economic implications of the society for assisted reproductive technology embryo transfer guidelines: Healthcare dollars saved by reducing iatrogenic triplets. *Fertility and Sterility*, 106(1):189–195.

Lemos, E. V., Zhang, D., Van Voorhis, B. J., and Hu, X. H. (2013). Healthcare expenses associated with multiple vs singleton pregnancies in the United States. *American Journal of Obstetrics and Gynecology*, 209(6):586–586.

Lundborg, B. P., Plug, E., Rasmussen, A. W., Lundborg, P., Plug, E., and Rasmussen, A. W. (2017). Can women have children and a career ? IV evidence from IVF treatments. *American Economic Review*, 107(6):1611–1637.

Machado, M. P. and Sanz-de-Galdeano, A. (2015). Coverage of infertility treatment and fertility outcomes. *SERIEs*, 6(4):407–439.

Martin, J. A. and Park, M. M. (1999). Trends in twin and triplet births: 1980-97. *National Vital Statistics Reports*, 47(24).

Merritt, T. A., Goldstein, M., Philips, R., Peverini, R., Iwakoshi, J., Rodriguez, A., and Oshiro, B. (2014). Impact of ART on pregnancies in California: An analysis of maternity outcomes and insights into the added burden of neonatal intensive care. *Journal of Perinatology*, 34(5):345–350.

Park, C. H. (2000). Prevalence of employer self-insured health benefits: National and state variation. *Medical Care Research and Review*, 57(3):340–360.

Reynolds, M. A., Schieve, L. A., Jeng, G., and Peterson, H. B. (2003). Does insurance coverage decrease the risk for multiple births associated with assisted reproductive technology? *Fertility and Sterility*, 80(1):16–23.

Schmidt, L. (2005). Infertility insurance mandates and fertility. *The American Economic Review, Papers and Proceedings*, 95(2):204–208.

Schmidt, L. (2007). Effects of infertility insurance mandates on fertility. *Journal of Health Economics*, 26(3):431–46.

Shapiro, C. L. and Recht, A. (2001). Side effects of adjuvant treatment of breast cancer. *New England Journal of Medicine*, 344(26):1997–2008.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76.

Tables

Table 1: Mandated infertility coverage in private health insurance plans

Generosity level	**State**	**Mandate year**	**IVF coverage**	**Notes**
0	Montana	1987	None	Applies to HMOs only; other insurers are specifically exempt from having to provide coverage.
0	New York	1990	None	Exempts coverage for IVF in the individual and small group markets, and coverage for GIFT or ZIFT. Must be 21-44 years old; must have the insurance policy at least 1 year before use; minimum 1 year of infertility if age $\leq$35 and min 6 month if age $>$ 35.
0	Ohio	1991	None	Does not define infertility. Requires HMOs to cover infertility services under basic health care services.
0	West Virginia	1995	None	Requires HMOs to cover infertility services under basic health care services.
1	Arkansas	1987	1 cycle	Lifetime $15,000 cap; minimum 2 years of infertility.
1	Hawaii	1989	1 cycle	Provides a one-time only benefit covering all outpatient expenses arising from IVF; minimum 5 years of infertility.
2	Connecticut	2005	2 cycles	Must be $<$40 years; mininimum 1 year of infertility; no more than 2 embryos implemented per cycle.
3	Maryland	1985	3 cycles	3 cycles per live birth, with a lifetime $100,000 cap. Businesses with $\leq$50 employees are exempt from mandated coverage.
3	Rhode Island	1989	3 cycles	Must be 24–40 years old; minimum 2 years infertility; $100,000 lifetime cap; insurer may impose up to a 20% co-payment.
4	Illinois	1991	4 cycles	Up to 4 egg retrievals; if a live birth occurs 2 additional egg retrievals covered for a lifetime maximum of 6 retrievals; minimum 1 year of infertility; Businesses with $\leq$25 employees are exempt.
4	New Jersey	2001	4 cycles	Min 2 years of infertility if age$\leq$35 and min 1 year of infertility if age$>$35; Must be $<$ 46 year.
5	Massachusetts	1987	No limit	No limit on the number of cycles or dollar lifetime cap; 1 year of infertility if age$\leq$35 and 6 month if age$>$35.

Notes: Source: RESOLVE: The National Infertility Association https://resolve.org [Accessed on November 2021].

Table 2: Summary statistics for Detail Natality Data, 1975-2014

	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
Multiple births per hundred live births	0.98	1.16	1.50	1.68	1.02	1.24	1.81	2.03
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of infants per thousand live births	1,009.92	1,011.89	1,015.55	1,017.24	1,010.38	1,012.74	1,018.96	1,012.90
Mean mothers' age	24.80	26.05	26.82	27.35	25.51	27.02	28.18	28.63
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mothers over 35 years (%)	4.39	7.79	11.74	12.86	5.55	10.16	16.49	18.16
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Married mothers (%)	82.16	73.24	66.52	60.08	78.80	72.31	67.98	61.69
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Mothers with college degree (%)	36.13	41.23	56.34	70.56	38.37	46.19	62.95	89.82
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
White mothers (%)	81.85	79.88	79.65	77.24	77.94	75.71	74.56	71.86
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
First time mothers (%)	36.54	32.94	33.21	32.44	36.38	33.61	32.18	31.33
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Number of births	17,578,332	19,207,128	19,849,815	20,966,038	5,009,715	5,701,859	5,477,201	5,217,796

Notes: Source: National Center for Health Statistics Detail Natality files. Weights constructed as described in Section 3 are used to calculate statistics in this table. Standard deviations appear in parentheses.

Table 3: Effects of mandated IVF coverage on multiple births per hundred live births, RDD model

	Connecticut		Rhode Island		New Jersey		Placebo New Jersey	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimated effect	-0.58***	-0.53***	-0.43***	-0.34***	-1.58***	-1.28***	0.0009	0.04
	(0.001)	(0.004)	(0.006)	(0.008)	(0.003)	(0.002)	(0.06)	(0.06)
Mean	7.95	7.95	5.72	5.72	23.06	23.06	7.71	7.71
	(27.05)	(27.05)	(23.23)	(23.23)	(42.13)	(42.13)	(26.68)	(26.68)
Cut off age	40	40	40	40	46	46	40	40
Bandwidth	1.94	1.95	1.97	2.01	2.43	2.24	3.26	3.63
Degree of polynomial	1	1	1	1	1	1	1	1
Covars	No	Yes	No	Yes	No	Yes	No	Yes
Number of observations	49,904	49,904	30,732	30,732	19,790	19,790	199,959	199,959

Note: This table presents the estimated effects of mandated IVF coverage on multiple births per hundred live births from the RDD model specified in Equation (1). The data includes all births to women ages 35 to 45 in Connecticut between 2007 (two years after mandated IVF coverage in 2005 for women below 40 years) and 2014, and in Rhode Island between 1991 (two years after mandated IVF coverage in 1989 for women below 40 years) and 2014, and all births to women ages 41 to 51 in New Jersey between 2003 (two years after mandated IVF coverage in 2001 for women below 46 years) from the birth certificate data. The running variable is women's age. The placebo estimates in New Jersey uses data on all the births to 35 to 45 years old women (an age window with no change in eligibility for IVF coverage) between 2003 (two years after mandated IVF coverage) and 2014. The included covariates are indicators for married, white, and college-educated women. The bandwidth and degree of the fitted polynomial are selected using Calonico et al. (2020). Standard errors are clustered at the age level and are presented in parentheses.

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

Table 4: Effects of IVF coverage generosity level on multiple births per hundred live births, GSC model

	All women		Women 35 and older		Women under 35		Number of state-year cells
	(1)	(2)	(3)	(4)	(5)	(6)	
A. All levels	0.10***	0.05***	0.18***	0.12**	0.05***	0.05	1,923
	(0.01)	(0.01)	(0.06)	(0.06)	(0.01)	(0.02)	
Pre-mandate mean	1.13	1.13	1.55	1.55	1,09	1.09	
	(0.30)	(0.30)	(0.57)	(0.57)	(0.23)	(0.23)	
B. Level 0	0.02	0.03	-0.22	0.01	0.05	0.02	1,404
	(0.05)	(0.04)	(0.17)	(0.13)	(0.07)	(0.03)	
Pre-mandate mean	1.05	1.05	1.43	1.43	1.02	1.02	
	(0.11)	(0.11)	(0.37)	(0.37)	(0.11)	(0.11)	
C. Level 1	-0.11*	0.08	-0.32*	-0.24	-0.10*	0.07*	1,110
	(0.06)	(0.04)	(0.15)	(0.16)	(0.05)	(0.04)	
Pre-mandate mean	0.96	0.96	1.39	1.39	0.93	0.93	
	(0.11)	(0.11)	(0.37)	(0.37)	(0.10)	(0.10)	
D. Level 2	0.16	0.16	0.23	0.15	0.02	0.24	936
	(0.11)	(0.13)	(0.27)	(0.27)	(0.06)	(0.11)	
Pre-mandate mean	1.46	1.46	2.09	2.09	1.33	1.33	
	(0.47)	(0.47)	(0.87)	(0.87)	(0.34)	(0.34)	
E. Level 3	0.17***	0.09*	0.52***	0.52***	0.12***	0.00	1,036
	(0.01)	(0.03)	(0.13)	(0.13)	(0.01)	(0.03)	
Pre-mandate mean	1.01	1.01	1.34	1.34	0.99	0.99	
	(0.08)	(0.08)	(0.37)	(0.37)	(0.07)	(0.07)	
F. Level 4	0.14***	0.17***	0.40**	0.31**	0.07**	0.12***	1,480
	(0.03)	(0.03)	(0.17)	(0.16)	(0.03)	(0.03)	
Pre-mandate mean	1.26	1.26	1.67	1.67	1.20	1.20	
	(0.33)	(0.33)	(0.62)	(0.62)	(0.25)	(0.25)	
G. Level 5	0.55**	0.28*	0.90**	0.56**	0.23**	0.21***	1,080
	(0.19)	(0.17)	(0.35)	(0.34)	(0.13)	(0.08)	
Pre-mandate mean	1.04	1.04	1.27	1.27	1.02	1.02	
	(0.08)	(0.08)	(0.14)	(0.14)	(0.08)	(0.08)	
Covars	No	Yes	No	Yes	No	Yes	

Notes: This table presents the estimated average treatment effect on the treated from the GSC model specified in Equation (2). Data aggregated to the state-year cell level. Included covariates in the model are mothers' age, marital status, education, and race; fathers' race; infant's sex; percentage of women of childbearing age; percentage of college-educated women; female labor force participation rate; the percentage of employees working in big firms (employee > 500); percentage with private health insurance; and real per capita income. Parametric bootstrapped standard errors estimated by 2,000 draws appear in parentheses.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 5: Effects of IVF coverage generosity level on the number of infants per thousand live births, GSC model

	All women		Women 35 and older		Women under 35		Number of state-year cells
	(1)	(2)	(3)	(4)	(5)	(6)	
A. All levels	1.06***	0.64***	1.64***	0.82**	0.52***	0.57***	1,923
	(0.08)	(0.07)	(0.30)	(0.42)	(0.07)	(0.12)	
Pre-mandate mean	1,011.61	1,011.61	1,015.95	1,015.95	1,011.12	1,011.12	
	(3.25)	(3.25)	(6.22)	(6.22)	(2.51)	(2.51)	
B. Level 0	0.18	0.27	-2.78	0.12	0.52	0.23	1,404
	(0.61)	(0.46)	(1.9)	(1.60)	(0.76)	(0.42)	
Pre-mandate mean	1,010.64	1,010.64	1,014.56	1,014.56	1,010.40	1,010.40	
	(1.22)	(1.22)	(3.85)	(3.85)	(1.15)	(1.15)	
C. Level 1	-1.25*	0.91***	-3.68***	-3.25*	-1.11*	0.77*	1,110
	(0.76)	(0.42)	(1.00)	(1.26)	(0.64)	(0.43)	
Pre-mandate mean	1,009.70	1,009.70	1,014.20	1,014.20	1,009.44	1,009.44	
	(1.12)	(1.12)	(3.71)	(3.71)	(1.07)	(1.07)	
D. Level 2	2.89	2.09*	4.30	2.09	1.53	2.39	936
	(1.50)	(1.29)	(2.57)	(2.62)	(0.99)	(1.18)	
Pre-mandate mean	1,015.13	1,015.13	1,021.88	1,021.88	1,013.74	1,013.74	
	(5.07)	(5.07)	(9.41)	(9.41)	(3.70)	(3.70)	
E. Level 3	1.94***	0.68***	3.73**	3.82***	1.34***	0.42	1,036
	(0.12)	(0.16)	(1.50)	(0.65)	(1.10)	(0.20)	
Pre-mandate mean	1,010.21	1,010.21	1,013.69	1,013.69	1,010.00	1,010.00	
	(0.78)	(0.78)	(3.85)	(3.85)	(0.76)	(0.76)	
F. Level 4	1.93***	1.66***	4.76***	3.43***	1.24***	1.31***	1,480
	(0.26)	(0.20)	(0.71)	(0.65)	(0.20)	(0.27)	
Pre-mandate mean	1,013.07	1,013.07	1,017.40	1,017.40	1,012.40	1,012.40	
	(3.71)	(3.71)	(6.95)	(6.95)	(2.80)	(2.80)	
G. Level 5	6.50**	2.92*	10.04**	6.35**	2.31**	2.18**	1,080
	(2.32)	(2.12)	(3.50)	(3.84)	(1.65)	(0.87)	
Pre-mandate mean	1,010.55	1,010.55	1,012.90	1,012.90	1,010.37	1,010.37	
	(0.84)	(0.84)	(1.48)	(1.48)	(0.79)	(0.79)	
Covars	No	Yes	No	Yes	No	Yes	

Note: See notes for Table 4.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 6: Effects of IVF coverage generosity level on patients' IVF utilization behavior, ME model

	All women				Women 35 and older				Women under 35	
	Total number of cycles		Average number of transferred embryos		Share of cycles		Average number of transferred embryos		Average number of transferred embryos	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
All levels	34.88***		-0.02		0.03***		-0.05		0.02	
	(13.11)		(0.05)		(0.01)		(0.06)		(0.04)	
Level 0		80.40**		0.07		-0.00		0.06		0.15
		(33.78)		(0.10)		(0.03)		(0.08)		(0.10)
Level 1		-34.79		0.01		0.06		-0.03		0.06
		(26.84)		(0.39)		(0.10)		(0.41)		(0.24)
Level 2		13.74		0.09***		0.03***		0.11***		0.06**
		(11.38)		(0.02)		(0.00)		(0.03)		(0.03)
Level 3		192.66***		0.00		0.09***		0.03		-0.07
		(37.71)		(0.06)		(0.01)		(0.05)		(0.08)
Level 4		42.06***		-0.05*		0.03***		-0.10***		0.02
		(13.27)		(0.03)		(0.00)		(0.04)		(0.04)
Level 5		650.22***		-0.50***		0.14***		-0.45***		-0.65***
		(14.27)		(0.04)		(0.01)		(0.04)		(0.04)
Constant	-372.10	-381.85*	4.32***	4.29***	0.44***	0.44***	4.42***	4.37***	4.35***	4.32***
	(230.14)	(204.30)	(0.52)	(0.52)	(0.10)	(0.10)	(0.51)	(0.50)	(0.59)	(0.59)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State and clinic random effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covars	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	4576	4,576	3,821	3,821	4,574	4,574	3,822	3,822	4,562	4,562

Notes: Source: SART's data of all women receiving IVF in a clinic in the US from 1996 to 2010. All estimates include year fixed effects and clinic random effects. Included state-level covariates from the CPS are listed in Notes to Table 4. We also control for the number of IVF clinics in each state. Standard errors are clustered at the state level and appear in parentheses.

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

Table 7: Effects of IVF coverage generosity level on adopted children per ten thousand live births, ME model

	All women		Women 35 and older		Women under 35	
	(1)	(2)	(3)	(4)	(5)	(6)
All levels	3.11		-11.83		-0.04	
	(6.79)		(52.77)		(1.23)	
Level 0		9.72		170.11		0.17
		(14.25)		(136.52)		(3.11)
Level 1		-29.76		-143.18		-5.36
		(19.23)		(208.76)		(5.66)
Level 2		16.57***		76.91**		1.33
		(6.26)		(39.23)		(1.65)
Level 3		5.88		-146.50		-1.15
		(24.27)		(153.16)		(3.12)
Level 4		-2.85		-65.27		0.92
		(7.73)		(55.73)		(1.24)
Level 5		-2.73		-233.43***		-2.11
		(7.59)		(62.94)		(1.75)
Constant	151.85	159.97	-53.39	-99.45	32.34	33.50
	(134.27)	(139.29)	(1340.20)	(1369.47)	(39.12)	(40.04)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State random effects	Yes	Yes	Yes	Yes	Yes	Yes
Covars	Yes	Yes	Yes	Yes	Yes	Yes
Observations	906	906	906	906	883	883

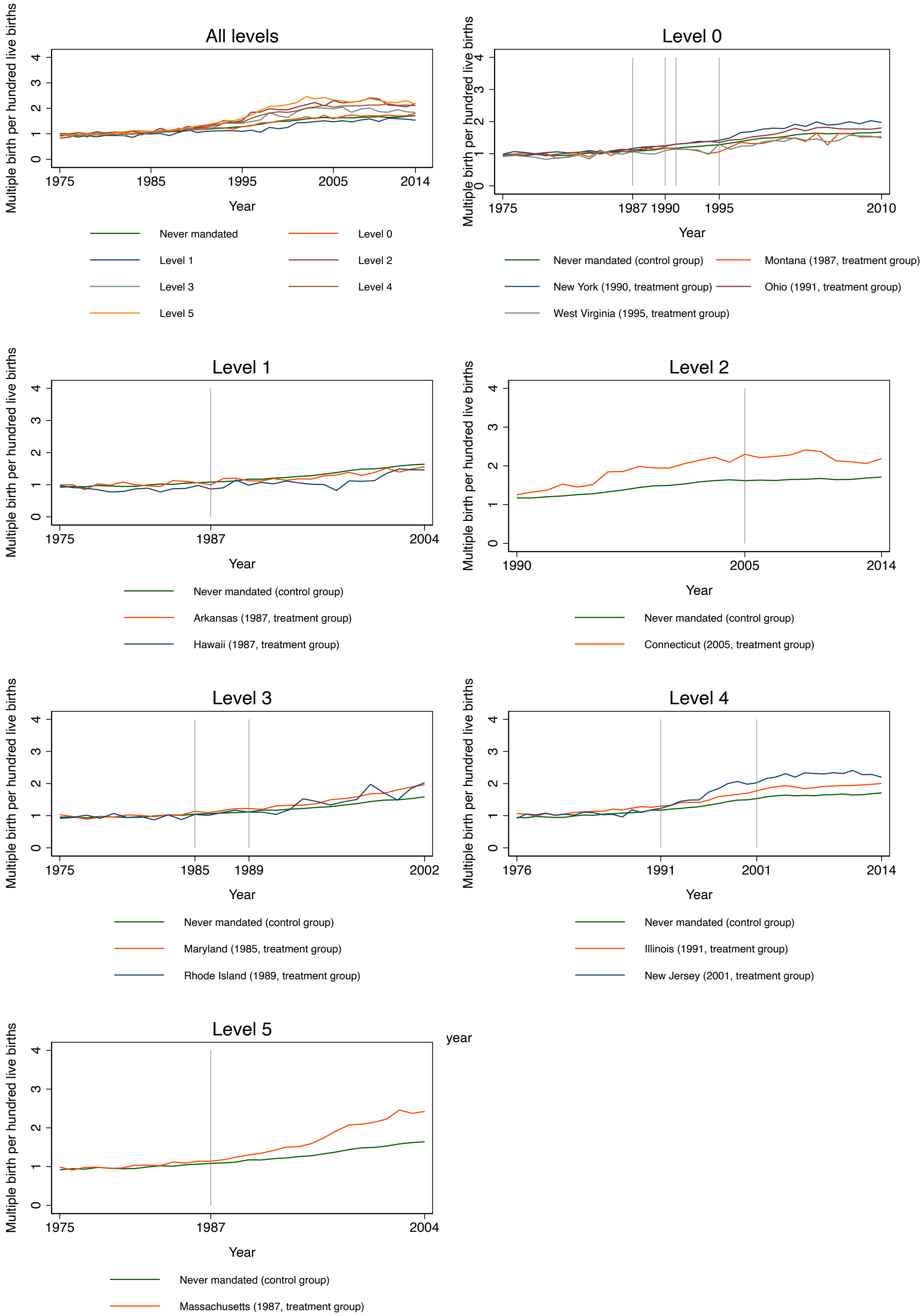
Note: Data include children ages 0-6 adopted between 1994 to 2014, aggregated into state-year cells. All estimated effects include year fixed effects and state random effects. Included state-level covariates from the CPS are listed in notes to Table 4. We also control for the number of IVF clinics in each state. Robust standard errors appear in parentheses.

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

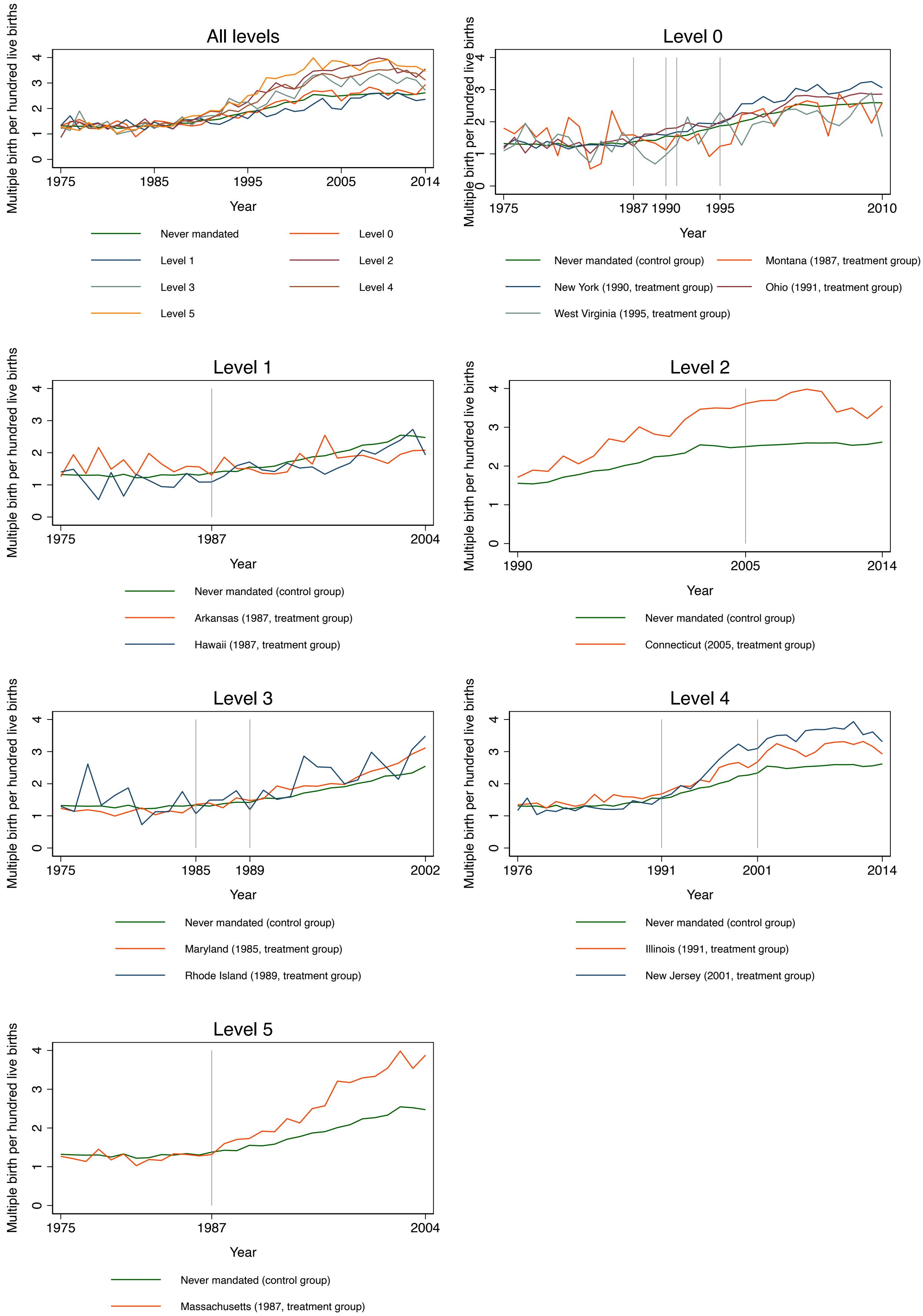
Figures

Figure 1: Multiple births per hundred live births by IVF coverage generosity level

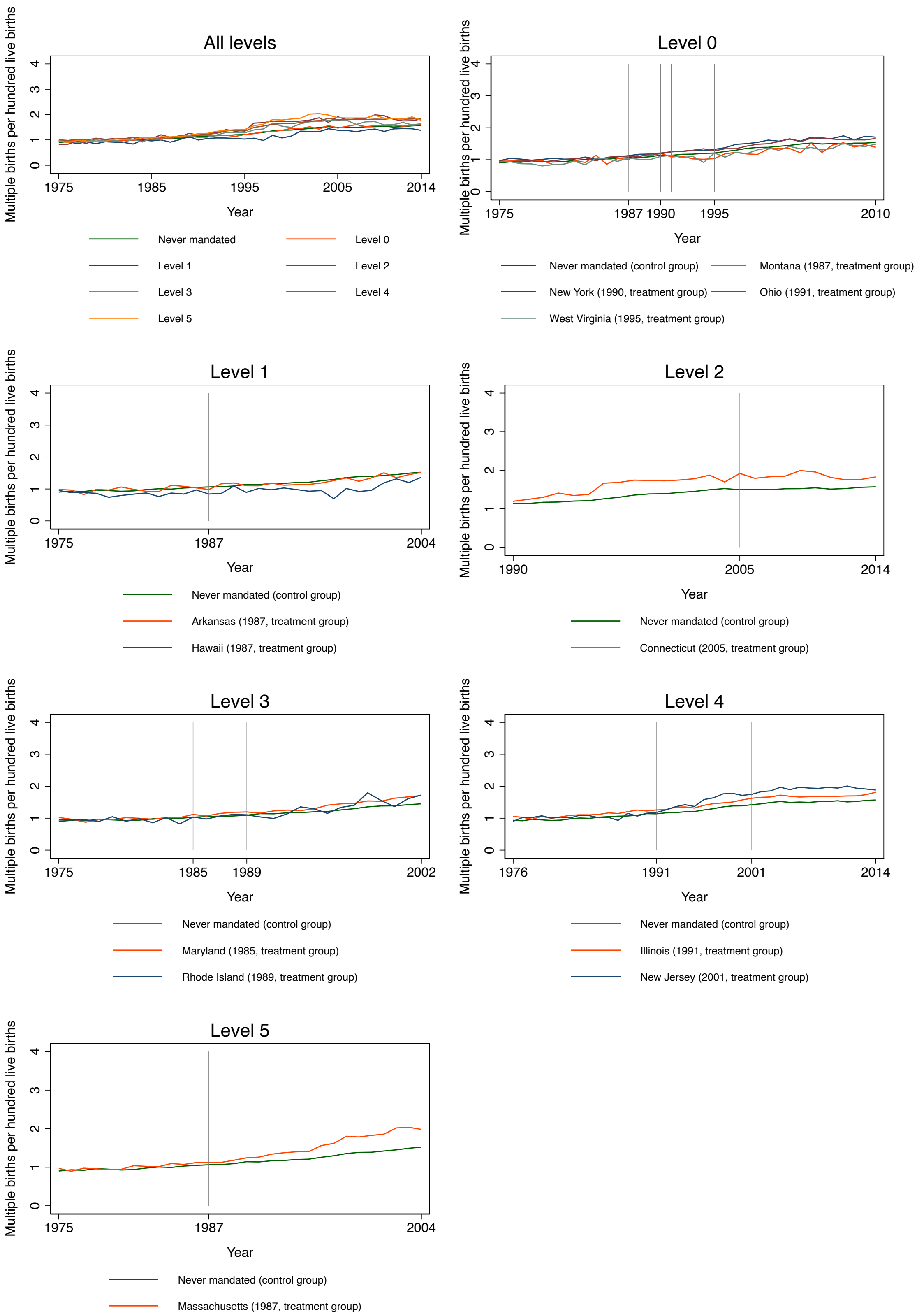
(a) All women



(b) Women 35 and older

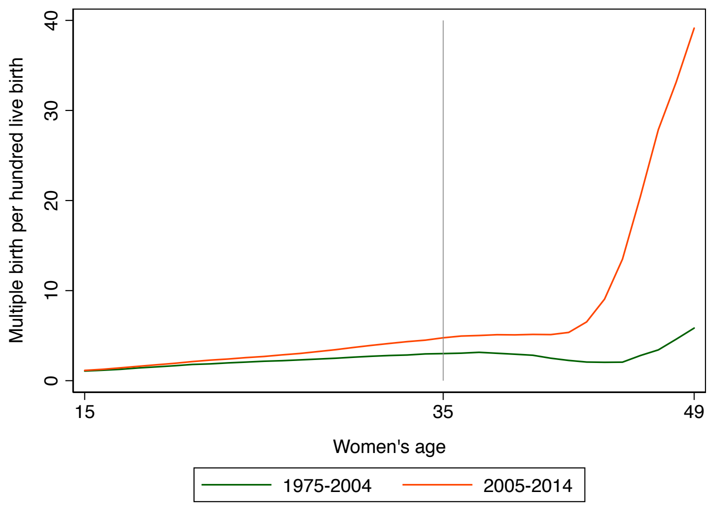


(c) Women under 35



Note: The sample includes all births from National Vital Statistics Detail Natality Data from 1975–2014. Multiple births are defined as births that are not singletons.

Figure 2: Multiple births per hundred live births by women's age

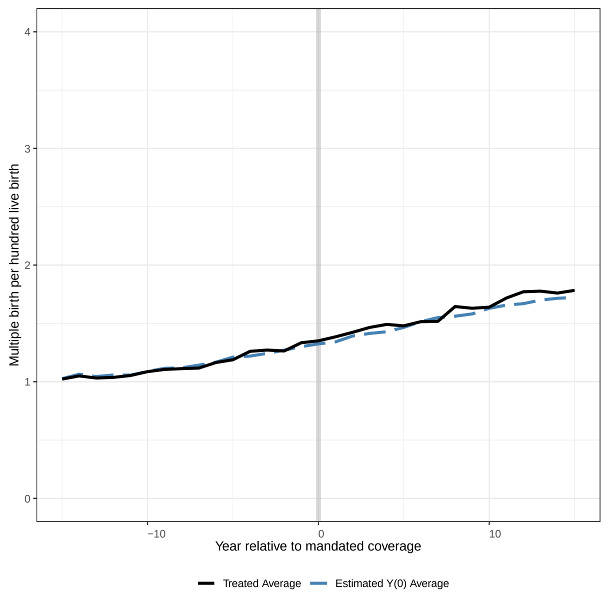


Note: Authors' calculations from the Detail Natality data. Multiple births are defined as births that are not singleton.

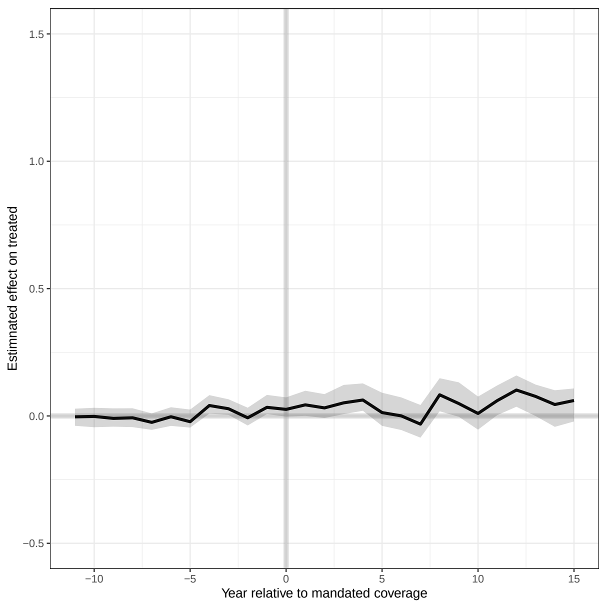
Figure 3: Effects of IVF coverage generosity level on multiple births per hundred live births, GSC model for all women

(a) All levels

(1) Treated average and estimated average for treated states

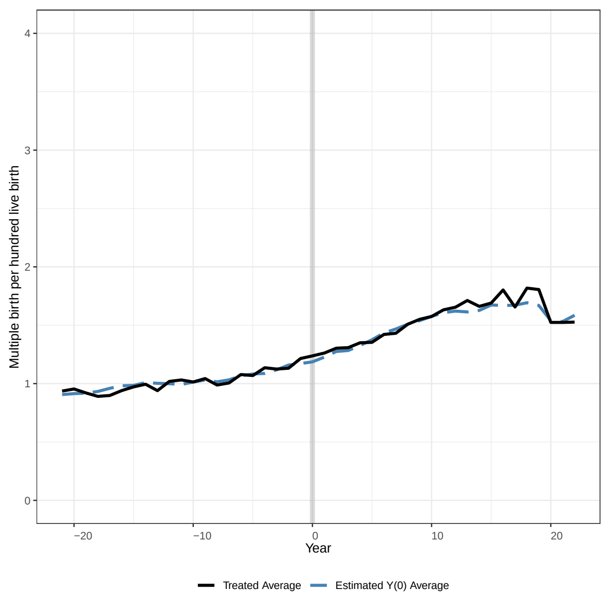


(2) Estimated treatment effect on treated

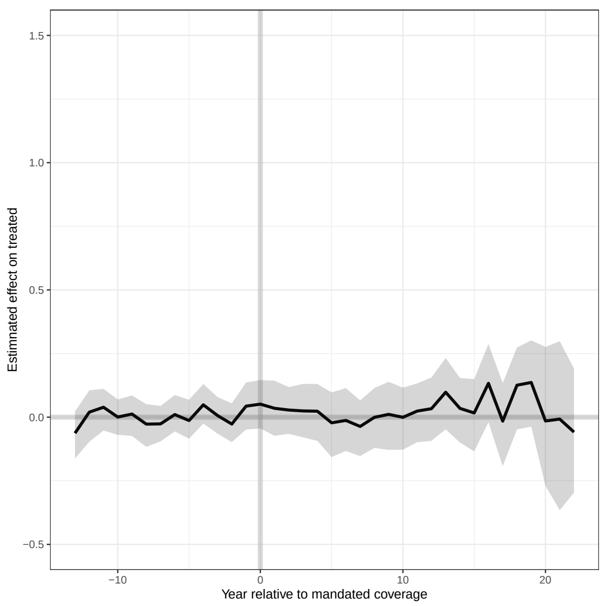


(b) Level 0

(1) Treated average and estimated average for treated states

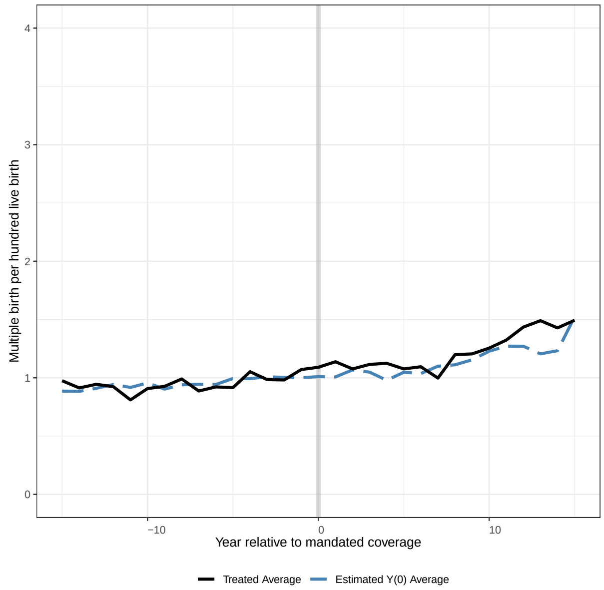


(2) Estimated treatment effect on treated

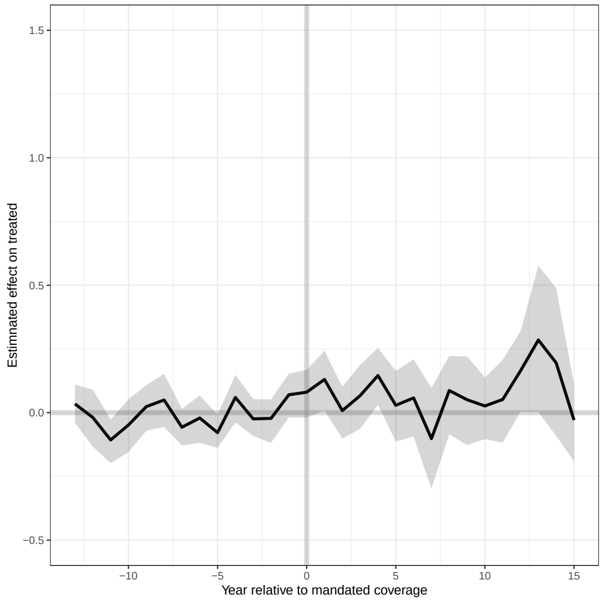


(c) Level 1

(1) Treated average and estimated average for treated states

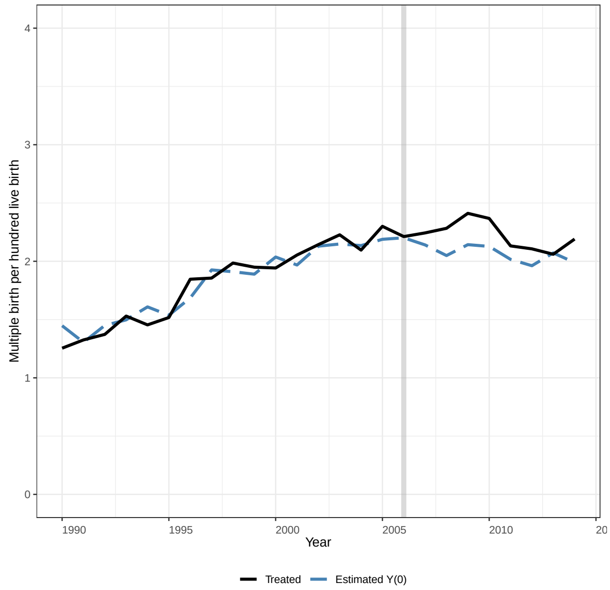


(2) Estimated treatment effect on treated

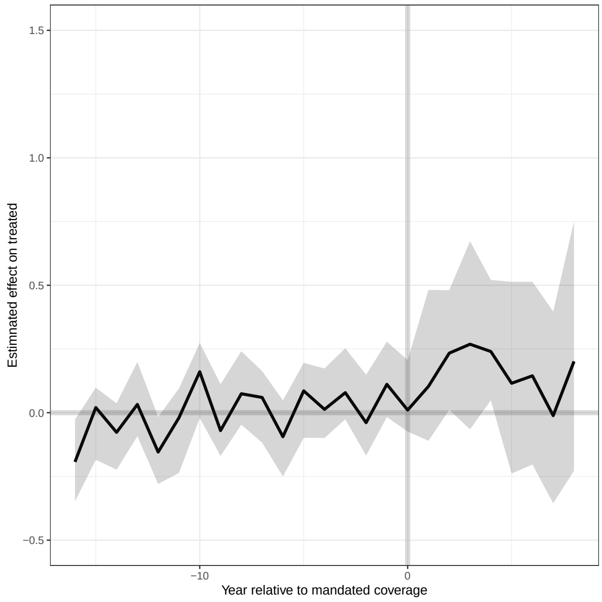


(d) Level 2

(1) Treated average and estimated average for treated states

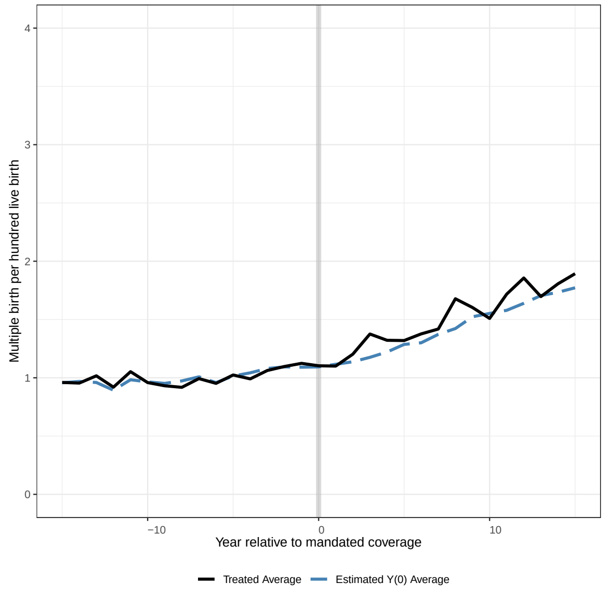


(2) Estimated treatment effect on treated

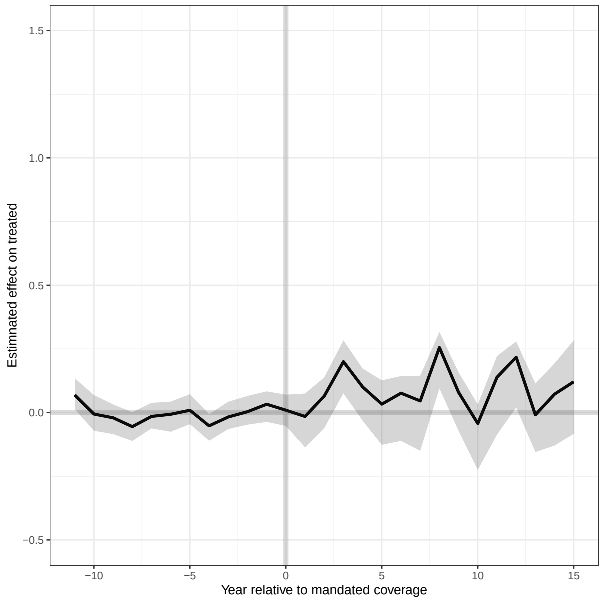


(e) Level 3

(1) Treated average and estimated average for treated states

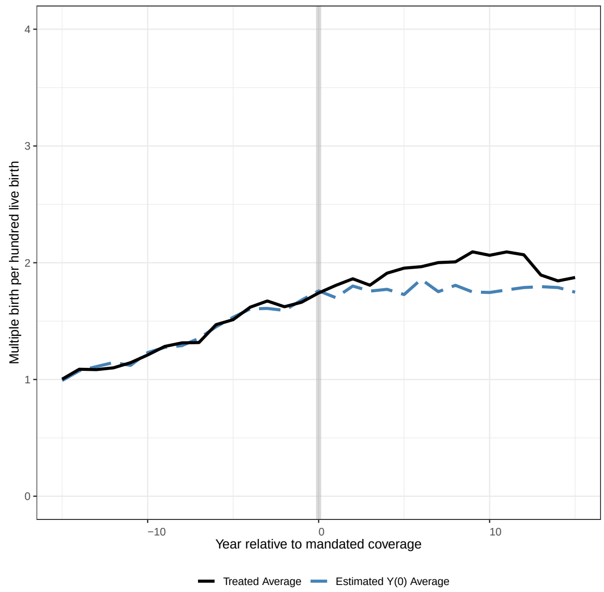


(2) Estimated treatment effect on treated

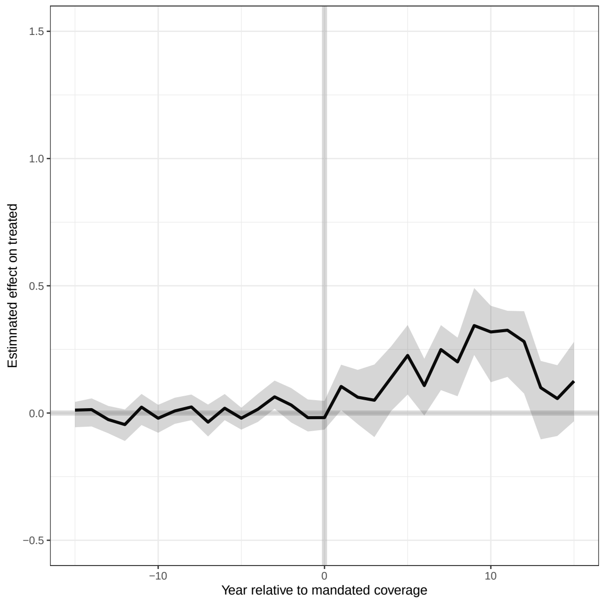


(f) Level 4

(1) Treated average and estimated average for treated states

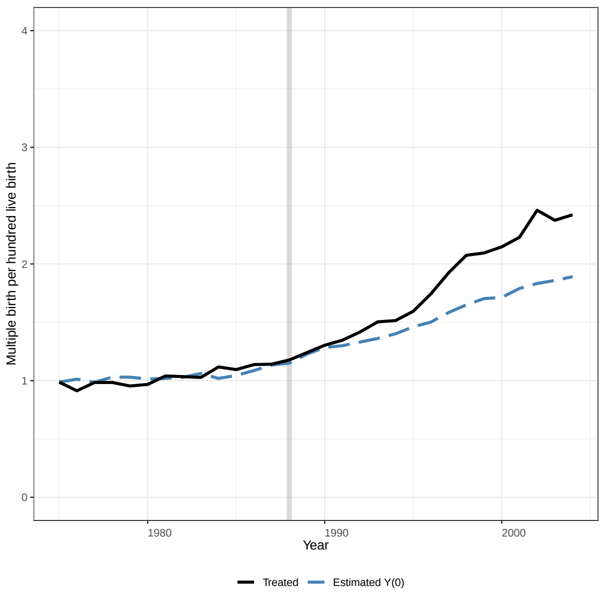


(2) Estimated treatment effect on treated

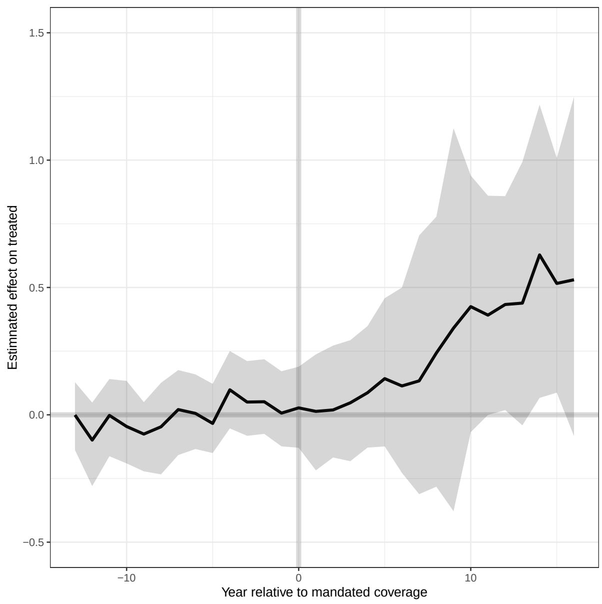


(g) Level 5

(1) Treated average and estimated average for treated states



(2) Estimated treatment effect on treated

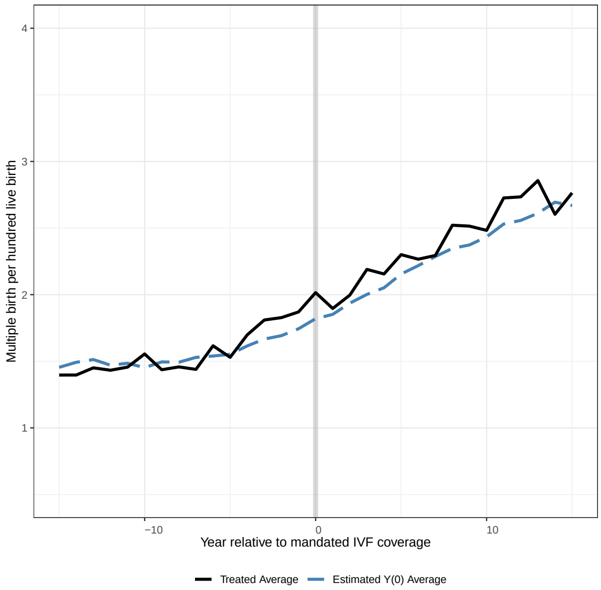


Note: This figure plots the estimated counter-factual outcome $Y(0)$ and the treatment effect on the treated on multiple births per hundred live births using the GSC model specified in Equation (2). The sample includes all births in the US from 1975-2014 from the National Vital Statistics, aggregated by state-year. The included covariates in the model are listed in the Notes to Table 4. The gray shade shows the %95 confidence intervals for the estimated effects.

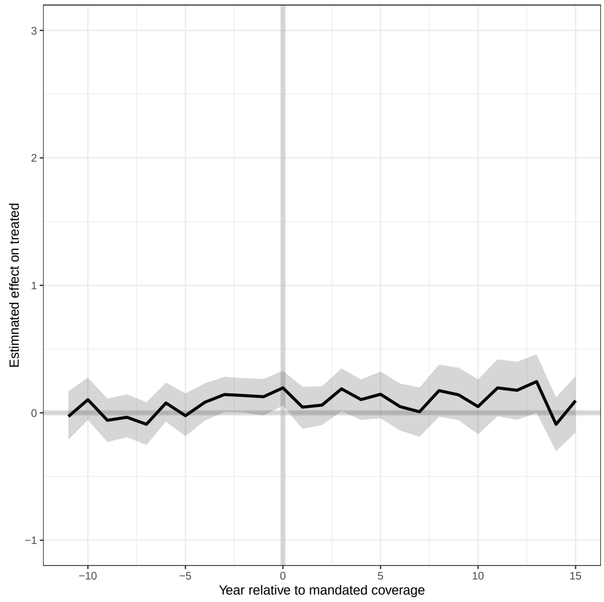
Figure 4: Effects of IVF coverage generosity level on multiple births per hundred live births, GSC model, women 35 and older

(a) All levels

(1) Treated average and estimated average for treated states

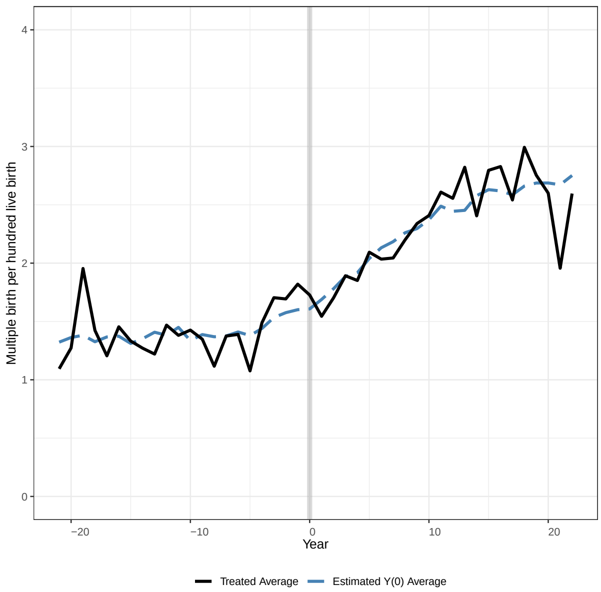


(2) Estimated treatment effect on treated

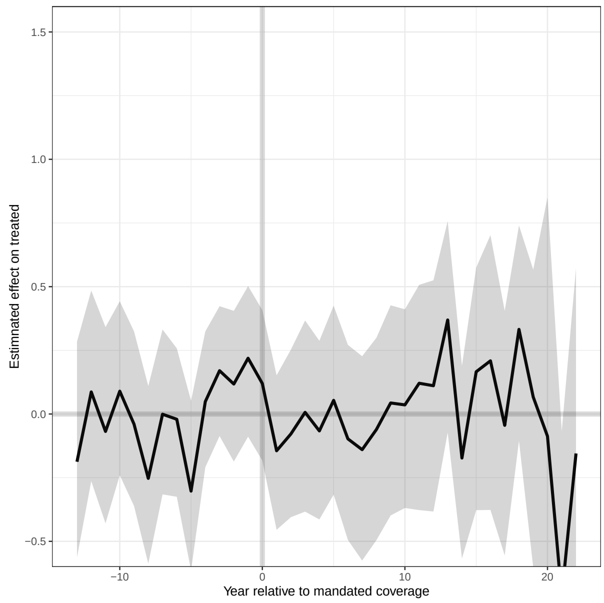


(b) Level 0

(1) Treated average and estimated average for treated states

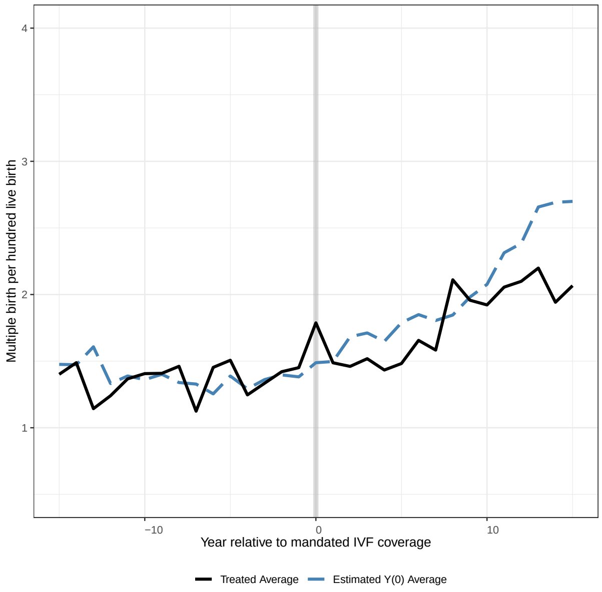


(2) Estimated treatment effect on treated

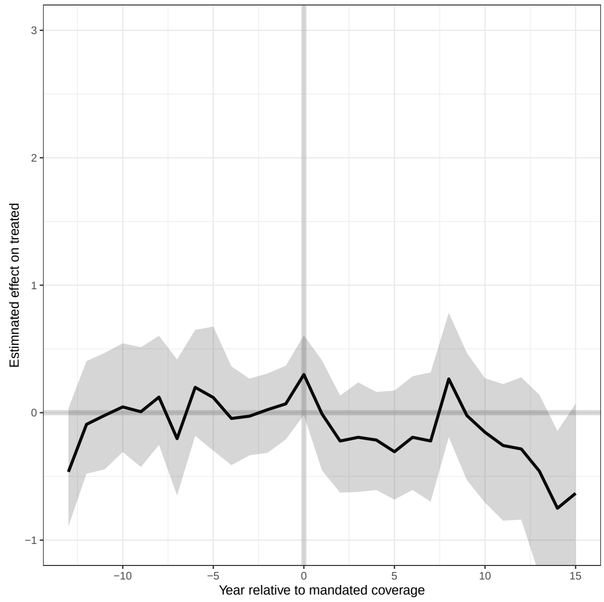


(c) Level 1

(1) Treated average and estimated average for treated states

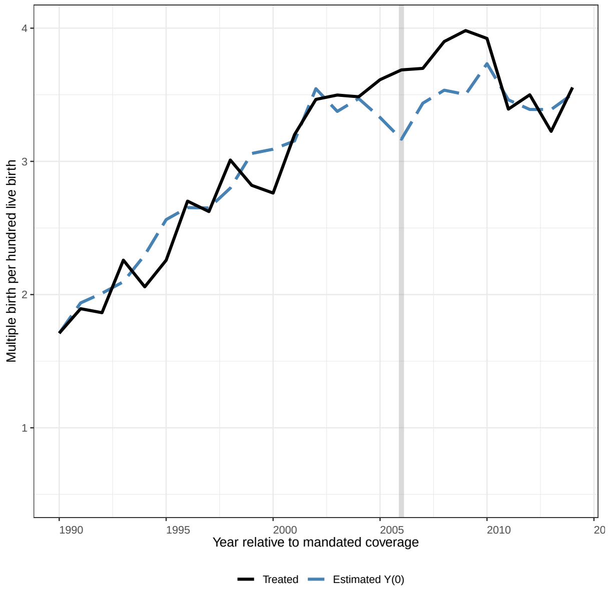


(2) Estimated treatment effect on treated

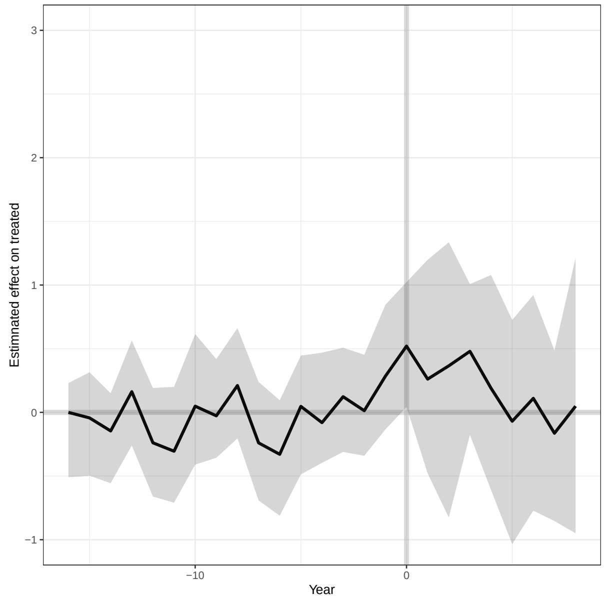


(d) Level 2

(1) Treated average and estimated average for treated states

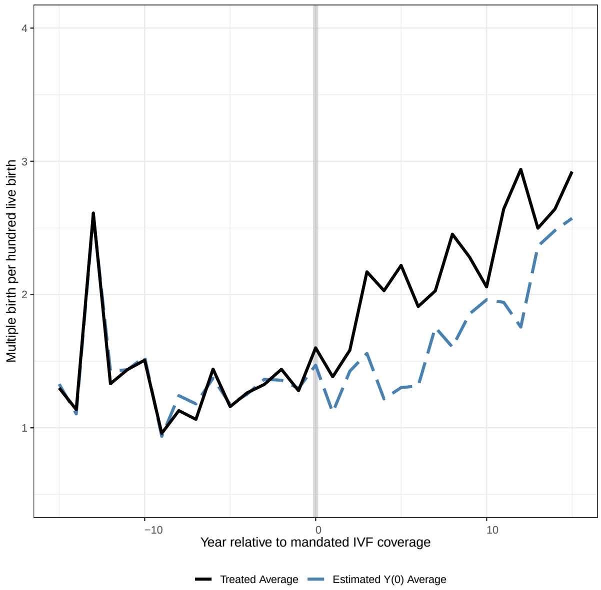


(2) Estimated treatment effect on treated

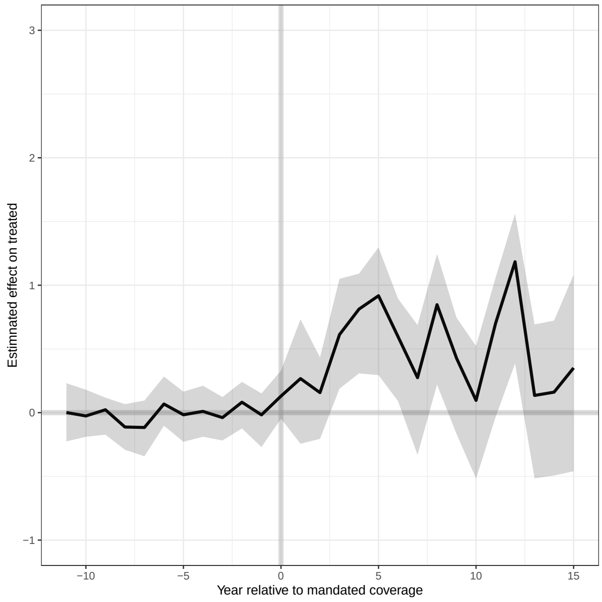


(e) Level 3

(1) Treated average and estimated average for treated states

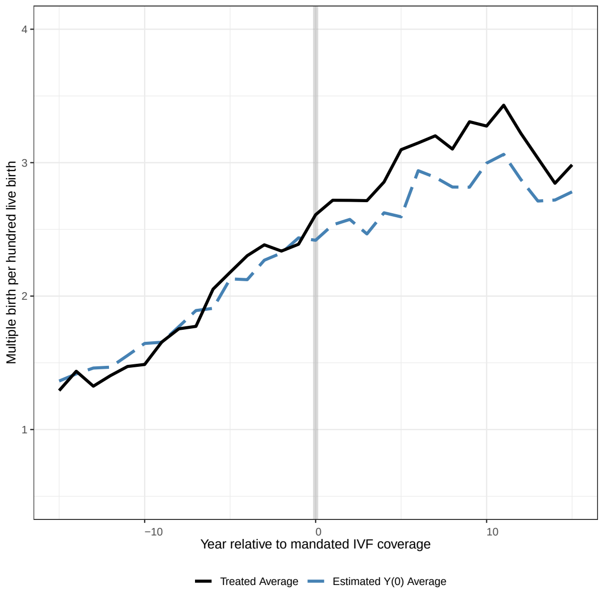


(2) Estimated treatment effect on treated

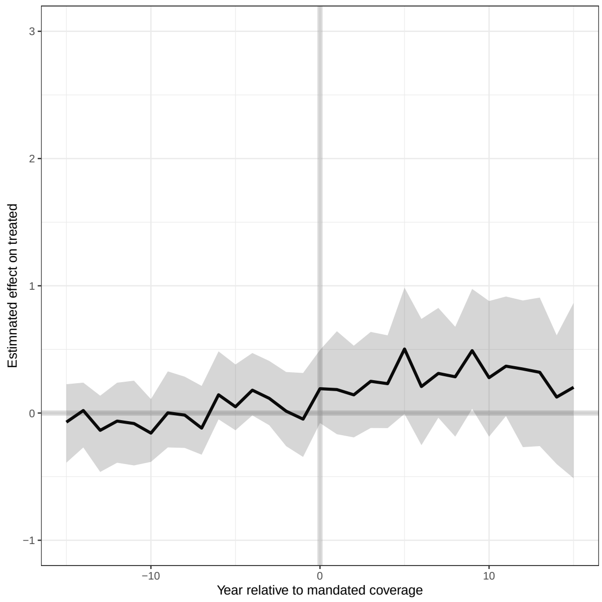


(f) Level 4

(1) Treated average and estimated average for treated states



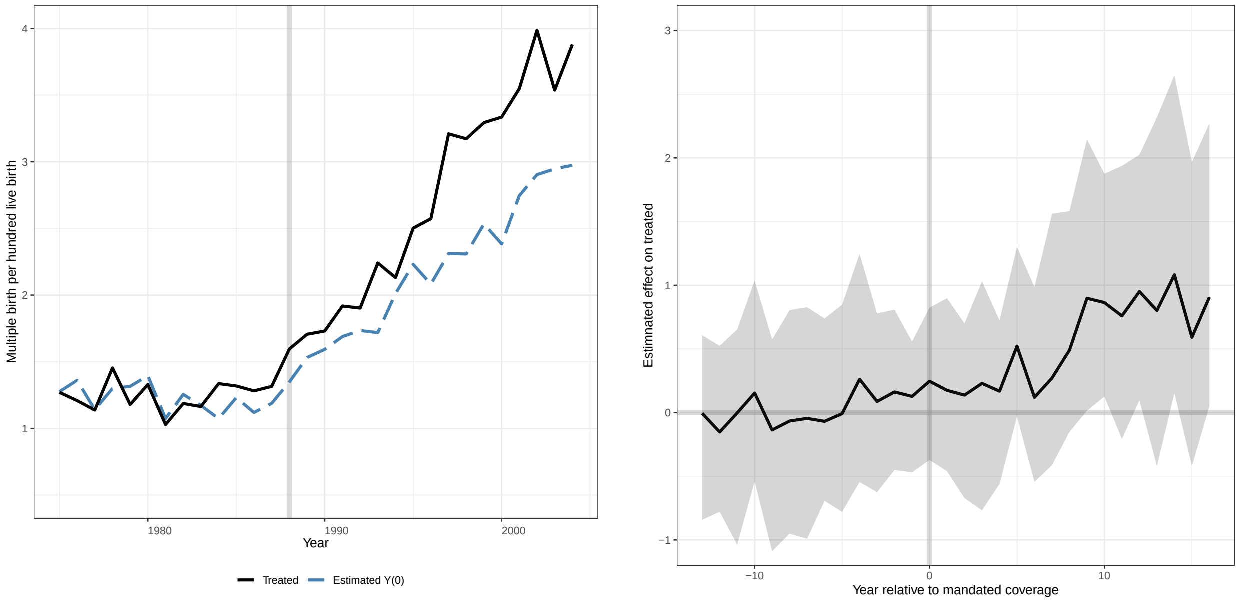
(2) Estimated treatment effect on treated



(g) Level 5

(1) Treated average and estimated average for treated states

(2) Estimated treatment effect on treated

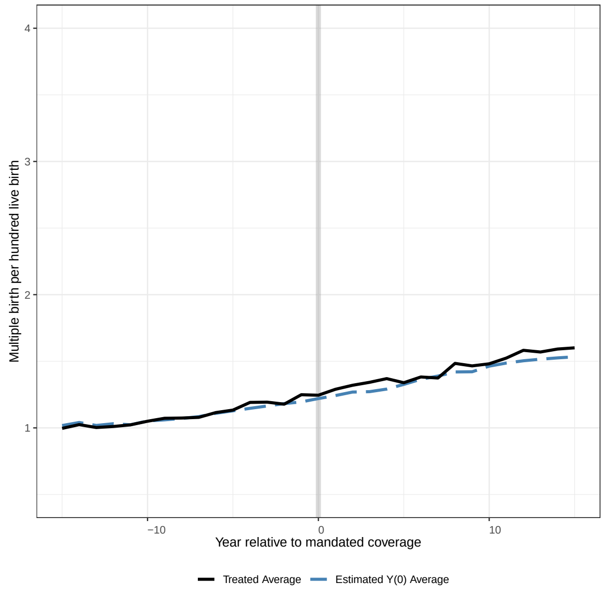


Notes: See notes for Figure 3.

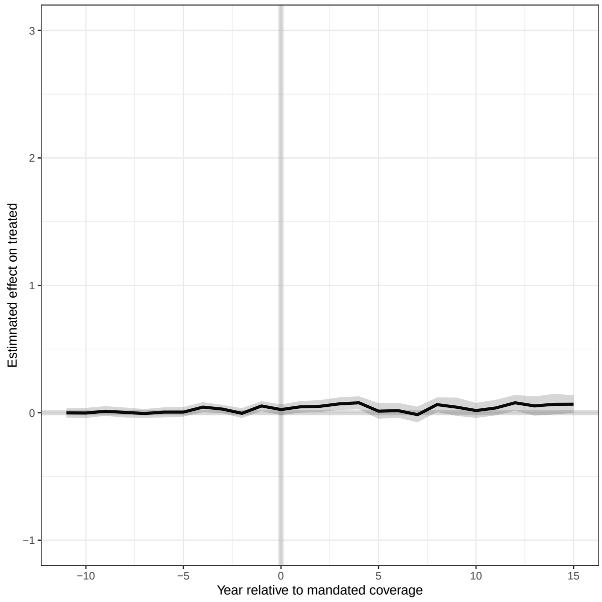
Figure 5: Effects of IVF coverage generosity on multiple births per hundred live births, GSC model, women under 35

(a) All levels

(1) Treated average and estimated average for treated states

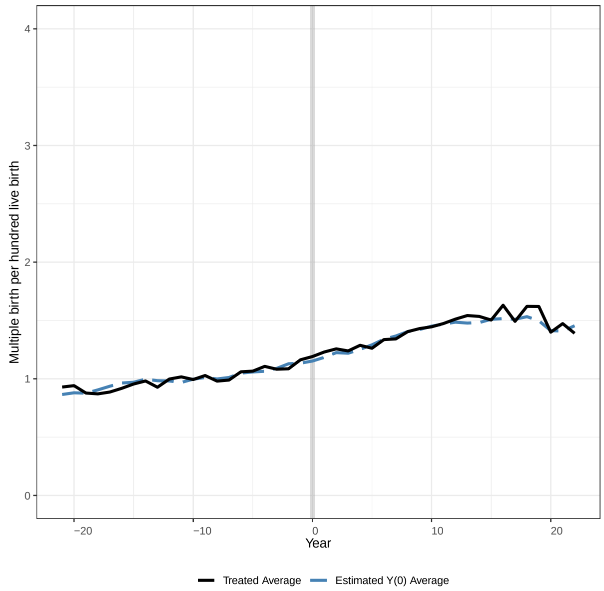


(2) Estimated treatment effect on treated

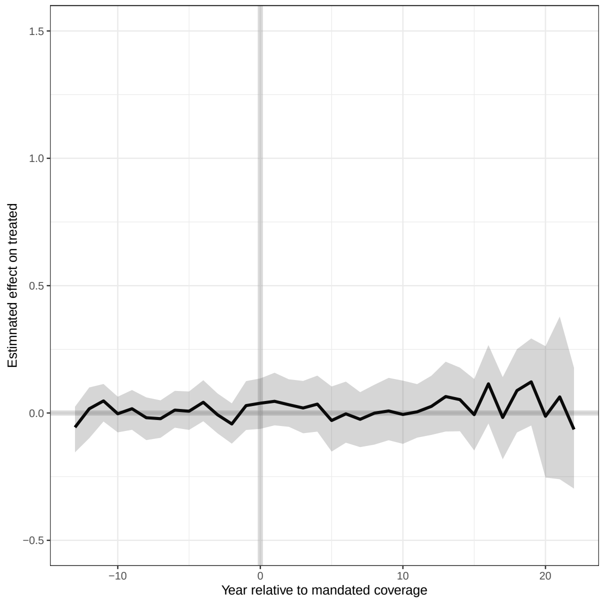


(b) Level 0

(1) Treated average and estimated average for treated states

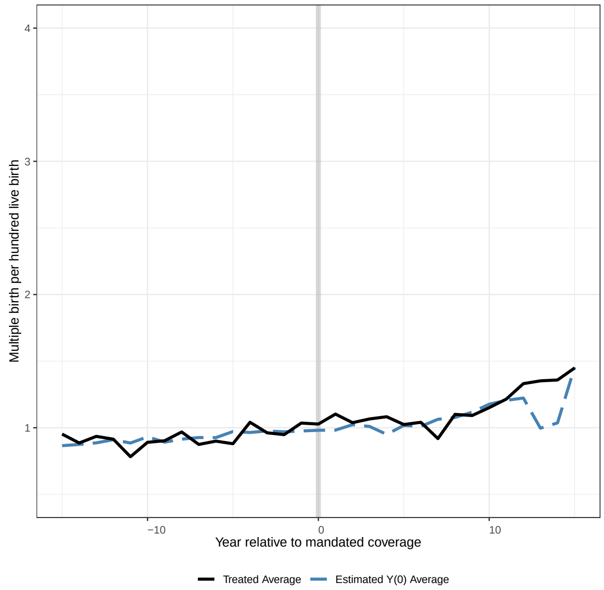


(2) Estimated treatment effect on treated

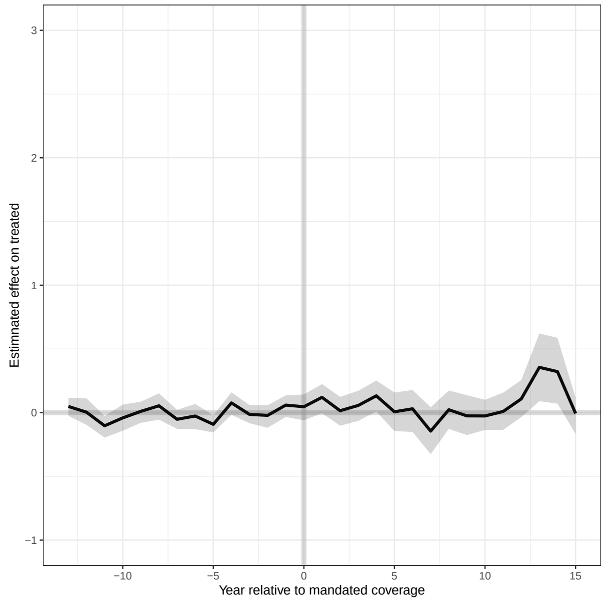


(c) Level 1

(1) Treated average and estimated average for treated states

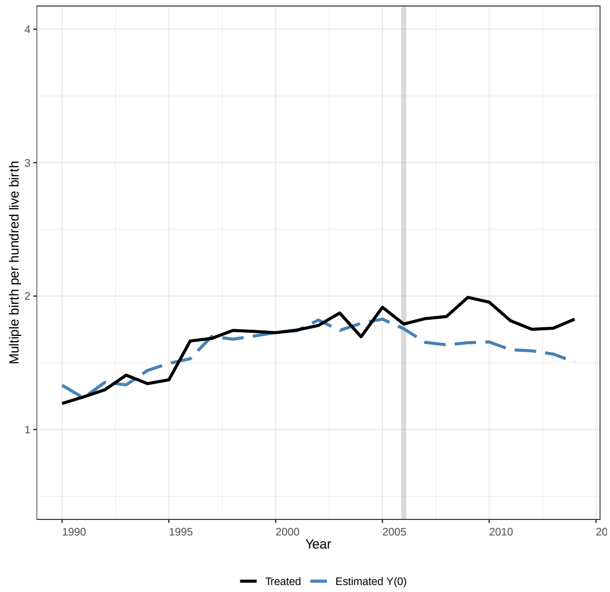


(2) Estimated treatment effect on treated

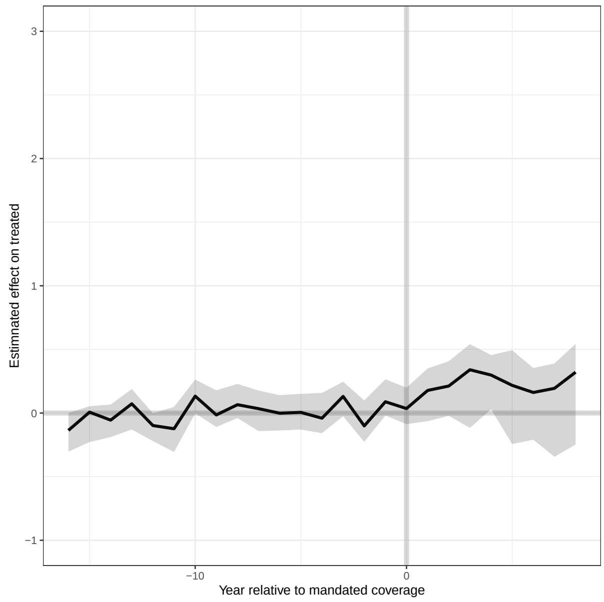


(d) Level 2

(1) Treated average and estimated average for treated states

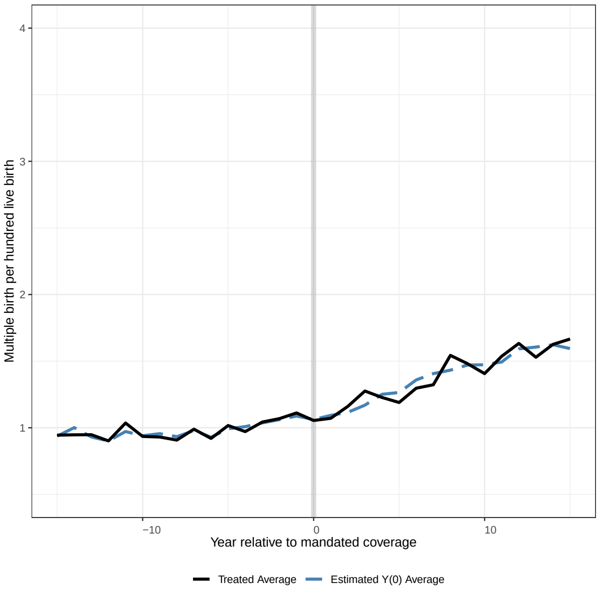


(2) Estimated treatment effect on treated

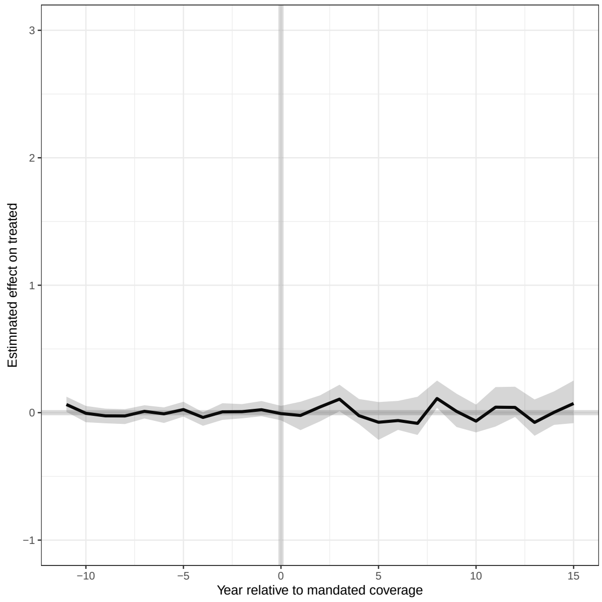


(e) Level 3

(1) Treated average and estimated average for treated states

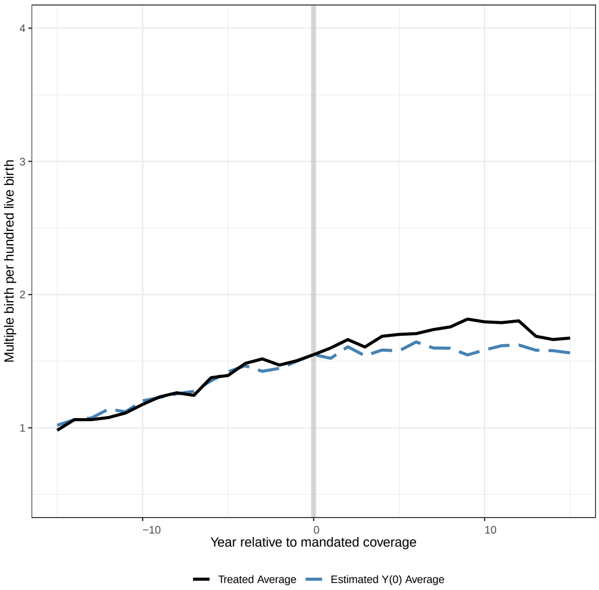


(2) Estimated treatment effect on treated

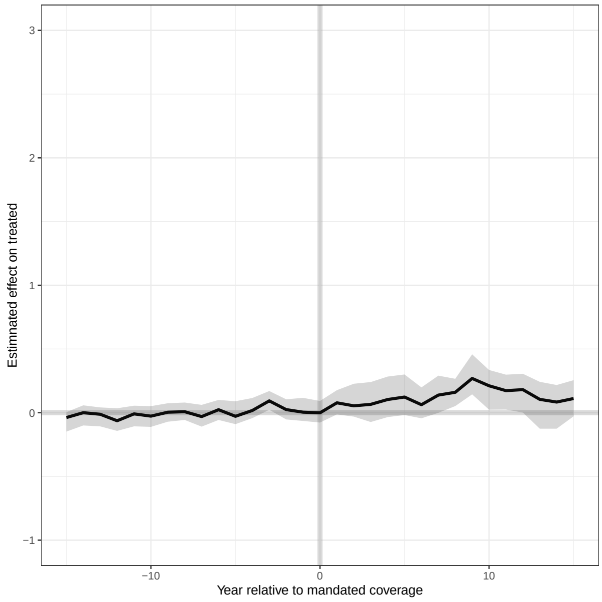


(f) Level 4

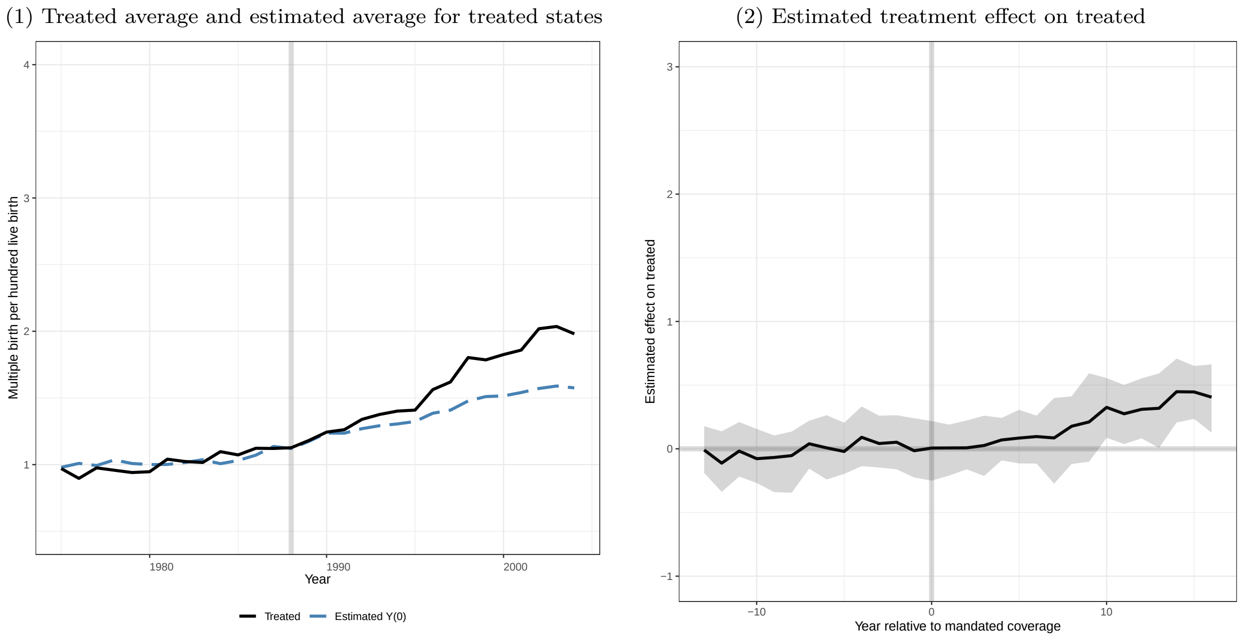
(1) Treated average and estimated average for treated states



(2) Estimated treatment effect on treated



(g) Level 5



Notes: See notes to Figure 3.

Appendix

A Summary statistics

Table A.1: Summary statistics for Society for Assisted Reproductive Technology (SART) infertility clinics 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	2.45	3.18	2.47
	(0.01)	(0.01)	(0.03)	(0.02)
Multiple births per hundred live births for all women	34.87	30.95	33.48	28.92
	(0.36)	(0.42)	(0.45)	(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	2.64	3.30	2.66
	(0.02)	(0.01)	(0.02)	(0.02)
Average number of embryos transferred for women under 35 years	3.12	2.23	3.03	2.22
	(0.02)	(0.01)	(0.03)	(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 (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 Error

[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).

(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 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:

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

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 respectively 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:

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

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 C.1 and Table C.2. The estimates in the first and second columns of each table 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.

[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 C.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
Covars	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
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: Study sample includes all births in the US from 1975-2014. Data aggregated into state-year cells for DD analysis and state-year-age cell for DDD analysis. All models include state- and year-fixed effects. Included covariates listed in notes for Table 4. Standard errors are clustered in state level and appear in parentheses.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table C.2: Effects of IVF coverage generosity level on the number of infants per thousand 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	1.07*	0.71***			1.72	1.22			0.59	0.57***			5.35***	2.75***		
	(0.55)	(0.26)			(1.23)	(0.89)			(0.39)	(0.20)			(0.95)	(0.93)		
Level 0			0.14	0.05			-0.45	-0.37			0.09	0.06			4.99***	2.57**
			(0.64)	(0.40)			(1.61)	(1.15)			(0.43)	(0.30)			(1.30)	(1.22)
Level 1			-1.26***	-0.20			-3.33**	-3.26***			-1.15***	-0.07			2.50	-1.22
			(0.18)	(0.30)			(1.58)	(1.19)			(0.39)	(0.23)			(2.22)	(1.33)
Level 2			1.38***	1.77***			3.79***	3.25***			0.33***	1.11***			6.29***	3.91***
			(0.12)	(0.27)			(0.31)	(0.66)			(0.10)	(0.22)			(0.00)	(0.38)
Level 3			2.23***	0.72			3.90***	3.30***			1.56***	0.65			6.31***	4.19***
			(0.36)	(0.54)			(0.37)	(1.12)			(0.25)	(0.42)			(0.59)	(0.75)
Level 4			2.33**	2.09***			4.60***	3.44***			1.38**	1.70***			8.19***	5.63***
			(0.90)	(0.74)			(1.37)	(0.92)			(0.54)	(0.50)			(0.31)	(0.32)
Level 5			4.60***	2.18***			9.03***	7.10***			2.96***	1.36***			10.28***	7.95***
			(0.17)	(0.47)			(0.38)	(1.01)			(0.15)	(0.38)			(0.00)	(0.44)
Constant	1010.11***	966.90***	1010.10***	984.36***	1013.55***	994.75***	1013.52***	1064.81***	1010.21***	979.11***	1010.20***	991.29***	1008.39***	1077.56***	1008.38***	1166.04***
	(0.16)	(23.51)	(0.16)	(26.30)	(0.65)	(79.62)	(0.66)	(79.25)	(0.15)	(21.29)	(0.14)	(24.97)	(0.38)	(61.43)	(0.38)	(35.75)
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
Covars	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
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: See notes for Table C.1. Standard errors are clustered in state level and are presented in parentheses.
$*p < 0.10, **p < 0.05, ***p < 0.01$

D Effects of the SART's guideline publication

SART uses a voluntary reporting system for IVF clinics to collect information on service utilization and outcomes, but it does not regulate clinics' practice. However, in January 1998, the SART published guidelines recommending the maximum number of embryos for transfer based on a woman's age and quality of embryos. These guidelines have been revised since then with more specified and restrictive changes in 2004. Table D.1 provides more information on the guidelines.

We use SART's clinic level data from 199 to 2010 to investigate the effects of the guideline published in 2004 on the average number of transferred embryos. We estimate an event study model specified as:

$$y_{ist} = \alpha + \rho Post_t \times Level_{is} + \beta X'_{ist} + \lambda_i + \lambda_t + \epsilon_{ist} \tag{D.1}$$

where i, s and t denote clinic, state and year respectively and y_{ist} denotes the outcome variable. We use the average number of transferred embryos as the outcome variable. $Post_t$ is a dummy switching on for years following the published guidelines.[3] $Level_{is}$ denotes the generously level of the mandated coverage. X_{st} denotes a set of time-varying state-level characteristics from the CPS described in the main text. λ_t and λ_i denote time and clinic fixed effects. ϵ_{ist} captures any remaining unobserved components affecting the outcome variable. The coefficient of interest is ρ, which captures changes in the average number of transferred embryos post publishing the guidelines compared to before publishing the guidelines.

Table D.2 presents the estimated effects. After publishing guidelines, the average number of transferred embryos is smaller, and the estimated effects are quite similar within all generosity levels and women's age.

[3] Similar to our previous analysis, we allow published guidelines in 2004 to affect multiple births with a two-year delay.

Table D.1: SART guidelines on the maximum number of embryos to transfer

Women's age (years)	1998	1999	2004	2006, 2008[a]	2009, 2013[a]
< 35 (*favourable*[b])	-	2	1-2	1-2	1-2
< 35	3	3	2	2	2
35-37 (*favourable*[b])	-	-	2	2	2
35-37	4	4	3	3	3
38-40 (*favourable*[b])	-	-	3-4	3	3
38-40	5	4	4	4	4
> 40 (*favourable*[b])	-	-	-	-	-
> 40	5	5	5	5	-
41-42 (*favourable*[b])	-	-	-	-	5
41-42	-	-	-	-	5

Note: This borrowed table from Lee et al. (2016) presents the SART's guidelines on the maximum number of embryos to transfer in an IVF cycle by women's age.
a In these years, guidelines were republished, but the recommended number of embryos to transfer per age group did not change.
b favorable: first cycle of IVF, good embryo quality, excess embryos available for cryopreservation, or previous successful IVF cycle.

Table D.2: Estimated effects from publishing SART guideline at 2004

	All women		Women over 35 years		Women under 35 years	
	(1)	(2)	(3)	(4)	(5)	(6)
All levels	-1.51***		-1.32***		-1.67***	
	(0.12)		(0.11)		(0.15)	
Level 0		-1.58***		-1.40***		-1.69***
		(0.16)		(0.15)		(0.22)
Level 1		-1.51***		-1.31***		-1.67***
		(0.10)		(0.09)		(0.14)
Level 2		-1.49***		-1.31***		-1.60***
		(0.13)		(0.12)		(0.18)
Level 3		-1.54***		-1.36***		-1.70***
		(0.11)		(0.11)		(0.15)
Level 4		-1.51***		-1.33***		-1.64***
		(0.14)		(0.13)		(0.18)
Level 5		-1.46***		-1.27***		-1.64***
		(0.10)		(0.09)		(0.13)
Constant	4.63***	4.78***	4.62***	4.77***	4.93***	4.87***
	(0.78)	(0.78)	(0.87)	(0.87)	(0.98)	(1.09)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Clinic fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Covars	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	1,520	1,520	1,520	1,520	1,723	1,723

Note: This table presents the estimated effects (ρ in Equation (D.1)) on the average number of transferred embryos in mandated to cover states form publishing SART guidelines in 2004. The study sample includes all clinics in mandated to cover states form SART's data from 1996–2010. See notes to Table 4 for a list of included covariates. Standard errors are clustered in state level and are presented in parenthesis.
$*p < 0.10, **p < 0.05, ***p < 0.01$