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

Arezou Zaresani[†] Lucie Schmidt[‡]

Prepared for Publication at the American Journal of Health Economics

Abstract

Mandated health insurance coverage for expensive in vitro fertilization (IVF) treatment varies widely in generosity across the US states. We find that more generous coverage within the states that mandate any coverage causes a greater incidence of multiple births, which are costly and can be risky. While more generosity is associated with fewer embryos transferred, this effect is outweighed by greater overall utilization of IVF. In addition, more generous coverage is associated with differences in the composition of patients, where more older women with lower fertility pursue treatment. This is mirrored by lower rates of child adoption by older women in those states. Utilization and 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, Anthony Scott, Hitoshi Shigeoka, Stefan Staubli, and Lindsay Tedds for their helpful comments. We have also benefited from discussions with conference and seminar participants at the University of Toronto (2022), University of Sydney (2022), 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). We also thank the editor of the American Journal of Health Economics, Professor Mireille Jacobson, and the two anonymous referees for their constructive comments. Maya Jasinska, Juliet May, and Elizabeth Van Arnam provided excellent research assistance.

[†]The University of Sydney, IZA and Tax and Transfer Policy Institute (TTPI) at Australian National University, Email: arezou.zaresani@sydney.edu.au. I acknowledge financial support from University of Calgary, University of Victoria, University of Melbourne, and University of Manitoba.

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

1 Introduction

Healthcare spending in the US has risen rapidly, from five percent of GDP in 1960 to 17.9 percent in 2017 (CMMS, 2017). Lifestyle changes and an aging population have contributed to increased 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 treatment. However, more generous coverage could also expand access to new patients, further contributing to increases in healthcare costs. This could be particularly important if more generous coverage changes 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 the ramifications of health policy interventions.

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 other a number of other dimensions, including age thresholds, coverage of unmarried women, and others (see Table 1). This variation allows us to identify the effects of coverage generosity on patients' 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).

In this paper, we empirically investigate how the generosity of mandated coverage for IVF treatment affects the overall incidence of multiple births, which are affected both by individual patient choices regarding the intensity of treatment as well as by the number and composition of patients 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.[1] More generous coverage could have 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 number and composition of 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.[2] 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 effects could dominate the decrease from the former, especially if the patients have a lower probability of success and transfer more embryos per cycle.

We first estimate a Regression Discontinuity Design (RDD) model, exploring age eligibility cutoffs in three mandate states to show that, in a causal sense, IVF utilization increases the incidence of multiple births.[3] We then use a Generalized Synthetic Control (GSC) model (Xu, 2017) to estimate the causal effects of IVF coverage generosity on the incidence of multiple births. We use 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 supplement our primary analyses with fertility clinic data from the Society for Assisted Reproductive Technologies (SART) from 1996 to 2010 to examine the association, in a descriptive sense, between more generous IVF coverage and the number of initiated IVF cycles, the composition of the pool of patients, and the number of embryos transferred per cycle. We also use data from the

[1]See Section 3 for a discussion of the weaknesses of this measure and section 4.2 for Regression Discontinuity Design (RDD) results that provide suggestive evidence that increases in multiple births are directly linked to coverage for IVF treatment.

[2]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. Abramowitz (2020) also discusses these intensive and extensive margin effects for the outcome of maternal mortality.

[3]As shown in Table 1, Connecticut, Rhode Island, and New Jersey have age restrictions.

National Data Archive on Child Abuse and Neglect (NDACAN) from 2000 to 2014 to examine the association between coverage generosity and child adoptions, which might be considered a substitute for conceiving through IVF.

Our paper is related to the literature investigating the effects of state infertiilty insurance mandates on a variety of outcomes, including utilization of treatment, infant health outcomes, fertility, age at first birth, maternal mortality, marriage timing, 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; Machado and Sanz-de-Galdeano, 2015; Abramowitz, 2017; Kroeger and La Mattina, 2017; Abramowitz, 2020).[4] 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]

Of these studies, the ones that relate most closely to our work look at multiple birth rates, utilization of IVF treatment, and the composition of those seeking treatment. On multiple births, most of the evidence suggests that mandates increase multiple births (Bundorf et al., 2007; Bitler, 2007; Buckles, 2013).[6] Studies that use fertility clinic-level data generally 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). In addition, studies find that mandates increase age at birth (Abramowitz, 2014, 2017; Kroeger and La Mattina, 2017) but do not increase total lifetime fertility (Machado

[4]A separate set of papers looks at IVF in other countries. Lundborg et al. (2017) uses IVF as an instrument for childbearing in the Danish context. Gershoni and Low (2021a) and Gershoni and Low (2021b) study the adoption of free IVF in Israel and find that it leads to later birth and fertility timing, as well as greater human capital investment and better labor market outcomes for women. Bhalotra et al. (2022) examines a Swedish policy of single embryo transfers and finds that it reduced the incidence of multiple births.

[5]One exception is Machado and Sanz-de-Galdeano (2015), which uses a synthetic control model to estimate the effects of mandated IVF coverage in the US on the timing of first births and women's total fertility rates.

[6]Buckles (2013) finds a small positive but insignificant effect on overall multiple births, but significant increases in triplet and higher births. A separate paper by Kulkarni et al. (2013) does not look at mandates but descriptively tries to analyze how much of the increase in multiple birth rates over time could be due to IVF.

and Sanz-de-Galdeano, 2015), which speaks to compositional effects of mandated IVF coverage.

Most of these papers in some way separate mandates that include IVF from those that do not, but only a few look at differences in generosity *within* the set of states that cover IVF. Several papers define "strong" versus "weak" mandates, with "strong" being defined as covering IVF and covering at least 35% of women (e.g., Buckles (2013); Machado and Sanz-de-Galdeano (2015)). Our work is closest to the papers that define "comprehensive" versus "limited" mandates, where "comprehensive" includes the early-adopting states that cover four or more cycles (Massachusetts, Illinois, Rhode Island) and "limited" includes all other states with some coverage (e.g., Jain et al. (2002); Reynolds et al. (2003); Bundorf et al. (2007); Henne and Bundorf (2008); Hamilton and McManus (2012)).[7]

Our main contribution to this literature is to examine more comprehensively how patients' utilization responds to the generosity of mandated IVF coverage. If policymakers wish to enact mandated coverage for IVF, one crucial element of policy design is the generosity of coverage. As discussed above, several papers look separately at the most comprehensive mandates, but by doing so, they miss potential differences in impacts between, for example, the least generous IVF mandate states and those that are somewhere in the middle. We further use clinic data to examine differences by generosity level in patients' utilization behaviors. We also provide evidence of the spillover effects from different generosity levels of mandated IVF coverage on child adoption as a close alternative to conceiving one's own infant.[8] Finally, our work benefits from methodological advances relative to the straightforward two-way fixed effects DD methods used in previous work.[9]

[7]New Jersey would also fall into this category, but its mandate was passed after the sample period used by these studies. Hamilton and McManus (2012) uses slightly different terminology, referring to these three states as having "universal" mandates.

[8]Previous studies have examined the relationship between IVF treatment and child adoption (Gumus and Lee, 2012; Cohen and Chen, 2010). However, the effects of mandated coverage on adoption could be heterogeneous, depending on the generosity of coverage and women's age.

[9]Our work also allows analysis of the more recent mandates legislated in the 2000s, which were not covered in much of the previous literature. This could be particularly important in the context of generosity since the studies looking at "comprehensive" mandates were only able to examine the three states of Massachusetts, Illinois, and Rhode Island, as the New Jersey mandate was passed after their data ended.

Four main findings emerge from our empirical analysis. First, we find a discontinuity in the incidence of multiple births at the age cutoff in mandate states with age restrictions, suggesting that IVF utilization is contributing to increases in multiple births. Ineligible women just above the cutoff experience a 5.5% to 7.3% decrease in the likelihood of a multiple birth compared to eligible women below the cutoff. Second, 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, the multiple birth rate increases by 27% relative to states with no mandated coverage, 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. Third, states with more generous coverage see a significantly larger number of cycles performed, but have fewer average transferred embryos per cycle. This is true for both older and younger women. Finally, states with more generous coverage have a significantly higher share of cycles performed on older women, suggesting that patients with lower fertility are being drawn into treatment. This is mirrored by a lower rate of child adoption by older women in states with more generous IVF coverage. Our findings suggest that the change in overall utilization outweighs the reduction in average embryos transferred per cycle, leading to overall increases in the incidence of multiple births. In addition, the compositional change towards more cycles to older mothers is likely to affect both economic costs and health risks.

Our findings suggest that changes in utilization and in the composition of the pool of patients are essential to understanding the policy implications of increased health insurance generosity. This is consistent with previous studies on the role of incentives in healthcare utilization. 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 intensive treatments could be optimal. Consistent with the work by Hamilton et al. (2018), Bhalotra et al.

(2022) find that a Swedish single embryo transfer policy reduced the incidence of multiple births and improved maternal and infant health.[10]

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 unsuccessful, assisted reproductive technologies such as IVF treatment are often recommended. Success rates of a single IVF cycle are as low as 20% (CDC, 2015), and many patients require more than one treatment cycle to achieve a live birth. One cycle of IVF treatment can cost as much as 46 percent of the average US family's annual disposable income (Kissin et al., 2016).

IVF treatment includes extracting eggs, obtaining a sperm sample, and manually combining eggs and sperm. The fertilized eggs, called embryos, are then transferred into the woman's uterus. The American Society of Reproductive Medicine practice committee provides guidelines on the maximum number of embryos to transfer per cycle (Klitzman, 2016).[11] However, given the high costs and low success rates of IVF, patients have the incentive to implant multiple embryos to increase their odds of success and, in doing so, increase the likelihood of multiple births. Most monetary costs of multiple births are covered by insurance, and many patients with fertility problems view multiple births as a desirable outcome (Gleicher and Barad, 2009; Barishansky et al., 2022), even though multiple births are costly and risky for both mothers and infants (Merritt et al., 2014;

[10] We are unable in our data to tell whether the increase in multiple births is due to women implanting more embryos than recommended by current medical practice. However, our SART data show that the average maximum number of embryos transferred across clinics is greater than SART guidelines would recommend. For example, in 2013-2015, the recommendation for women ages 38-40 is three embryos, but the average maximum across clinics is five.

[11] Currently, recommendations are for 1-2 embryos per cycle for women under 35 years old and increase with age.

Caserta et al., 2014).[12]

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 twelve states in the US passed legislation pertaining to coverage of infertility treatment in employer-provided private health insurance plans.[13] In these *mandate to cover* states, private health insurance companies are required to cover infertility treatment in all of their policies.[14]

The level of coverage in the *mandate to cover* states is quite heterogeneous. Some state mandates do not require any coverage of IVF. Within the set of states that do, mandates vary along a number of dimensions. Most important is the number of covered cycles, but mandates vary along other dimensions as well (see Table 1).[15] However, these dimensions of generosity are generally time invariant once a mandate is passed, and are highly correlated with the mandated number of cycles covered, so we treat the number of cycles as a proxy for the overall generosity level of mandated coverage. During our study period, Montana, New York, Ohio, and West Virginia mandate some coverage for fertility treatment but do not require coverage of IVF. We group these states as Level 0 coverage. Arkansas and Hawaii have the least generous coverage, and we group them as Level 1. Connecticut is the only state with Level 2 coverage. Rhode Island and Maryland have Level 3 coverage, and Illinois and New Jersey are grouped as states with Level 4 coverage. Massachusetts, with an unlimited number of covered cycles, has the most generous mandate and has Level 5 coverage. There are 35 states that do not mandate

[12]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).

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

[14]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.

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

coverage for infertility treatments.[16] These *never mandate* states serve as a control group in our analysis. As a robustness test below, we examine whether other dimensions of coverage besides the number of cycles affect the relationship between mandated coverage and multiple births.

3 Data

We use several data sources for our empirical analysis. First, we use 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 the mother's age, education, and race, the father's race, parental marital status, and state of residence, and infant information such as sex, birth order, and plurality (single or multiple births). Our study sample includes the twelve 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.[17]

Our primary outcome variable is the multiple birth rate, defined as the number of multiple births (i.e., not singletons) per hundred live births.[18] Multiple births are a valuable proxy for the intensity of treatment, as 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 multiple births are naturally occurring, due to IVF treatment, or due to other infertility treatment besides IVF.[19] Our simple multiple birth indicator also does not differentiate

[16] Since the end of our study period; four additional states have mandated IVF coverage: Colorado (2020), New Hampshire (2020), New York (2020), and Delaware (2018). These states are considered control states in our analysis.

[17] The public-use birth certificate data includes 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 for 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.

[18] 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.

[19] The birth certificate data includes a variable indicating births with assisted reproductive technology

between, for example, a twin birth and a quadruplet birth, even though these have very different cost implications, so we also examine the effects of generosity on the number of infants per thousand live births as an alternative outcome variable.

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 percentage of women of childbearing age, the female labor force participation rate, and real per capita income.[20] 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 Act (ERISA).[21]

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.[22] 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 per cycle. 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.[23] The data include the records of all the public adoptions in the US and has information on adoptive parents' age and race; the adopted children's age, sex, and race; and the year and the state in which the adoption is finalized. We focus on children from birth to age six since younger children might be closer substitutes for newborn infants, but we examine adoptions of older children as well.

starting from 2011. However, the variable has many missing values and is not very informative.

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

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

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

[23]The data are 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 early mandates in the 1980s and early 1990s.

We create a variable representing the number of adopted children per one thousand live births in each state-year cell.[24]

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. The 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.

Figure 1 plots trends in multiple births per hundred live births by generosity level of mandated IVF coverage separately for older and younger women.[25] Three main patterns emerge. First, the incidence of multiple births is increasing across all states over our study period. Second, more generous coverage is generally associated with more rapid growth in the incidence of multiple births. Third, the incidence of multiple births is higher for older women.[26]

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

Older women are more likely to have multiple births, even in the absence of infertility treatment (Hazel et al., 2020; Adashi and Gutman, 2018; Beemsterboer et al., 2006).[27]

[24] 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.

[25] 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.

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

[27] Women are more likely to conceive fraternal twins once they reach their 30s as a result of an evolutionary response to combat declining embryo viability (Hazel et al., 2020). Adashi and Gutman (2018) find that by the time white women reach age 35, they are about three times more likely to have fraternal, non-identical twins. African American women are four times more likely to have twins at age 35. The risk for triplets and quadruplets goes up four and a half times and six and a half times, respectively.

Figure 2 plots multiple birth rates by women's age and shows that the increasing pattern is stronger in recent years. To examine the extent to which IVF coverage might be responsible, we compare multiple birth rates of women eligible for mandated IVF coverage with those 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). We explore the sharp discontinuity in age eligibility for 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 without treatment, so $f(a)$ denotes the age trend to control for this relationship. 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 likelihood of a multiple birth. 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).[28]

We follow Schmidt (2007) and allow mandated coverage to affect multiple births with a two-year delay. This 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

[28] We do not directly examine the impact of the generosity of mandated IVF coverage on the total number of births, but previous papers have found that mandates increase the number of births, particularly for older women (e.g., Schmidt (2005, 2007)). If the mandates are also increasing the denominator, this would bias us against finding an effect on multiple births at the cutoff.

after the mandate date in each state up to 2014 and focus on women within the five-year window around the eligibility age for estimating Equation (1).[29]

Figure 3 presents the RDD plots, and Table 3 presents the estimated effects. 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. After controlling for individual characteristics, the incidence of multiple births for women just above the age cutoff (and therefore not eligible for mandated coverage) is 7.29%, 5.94%, and 5.55% (0.53, 0.34, and 1.28 percentage points) lower respectively in Connecticut, Rhode Island, and New Jersey compared to eligible women just below the age cutoff.[30] We then estimate the effects on the incidence of multiple births in New Jersey from a placebo eligibility age threshold of 40 years.[31] The last panel of Table 3 shows that the estimated effect is negligible and insignificant, providing additional confidence that the RDD estimates presented in the earlier columns are picking up causal effects. 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 next use a Generalized Synthetic Control (GSC) framework developed by Xu (2017) to estimate the impact of coverage generosity on the incidence of multiple births. Unlike the RDD model, which estimates local average treatment effects narrowly focused around the age eligibility cut-offs in only three states, the GSC model allows us to estimate causal

[29]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; 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.

[30]Sample means presented in Table 3 indicate that the multiple birth rate for our New Jersey sample is much higher than that for the relatively younger women in our Connecticut and Rhode Island samples. The sample mean for the placebo sample in New Jersey, where the women are of the same ages as women in the Connecticut and Rhode Island samples, is more similar to those sample means. This observation is consistent with the graphical evidence in Figure 2, which shows that multiple birth rates rise steeply for older women.

[31]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.

effects of the mandates for the full sample of women. Much of the previous literature on infertility mandates uses 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 generosity of mandates, and our GSC framework helps with this issue. In addition, recent methodological advances highlight problems with the standard DD or two-way fixed effects models when policy variation occurs in different states at different times (e.g., Goodman-Bacon (2021), de Chaisemartin and D'Haultfoeuille (2020)). Our analysis does not suffer from these issues since we compare states with a certain level of coverage (treatment group) to never mandated states (control group), so we are not contaminating our control group with previously treated states.[32]

The GSC model is a generalization of conventional synthetic control models using 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).[33] A GSC model uses the control and treatment groups (in pre-treatment periods) to impute treated counter-

[32]The mandate dates of states grouped in each level of coverage are generally quite close to each other (see Table 1), further limiting the contamination due to staggered adoption of mandated coverage. If there is variation in the treatment effect across states and time, then the estimated effects from a DD model would be a non-convex weighted average of the estimated effect for each state, where the weights sum to one but may be negative. The possibility of negative weights is concerning because, for instance, the treatment effect for each state could be positive (negative), and yet the estimated coefficient from a DD or two-way fixed effect model might be negative (positive) (see Roth et al. (2022) for more details). Figure E.1 in Appendix E plots the estimated weight of each mandated state over the years, which are the residuals from a regression of a mandated coverage indicator on state and year fixed effects, scaled by the sum of the squared residuals across a pooled sample of mandated and never mandated states (see de Chaisemartin and D'Haultfoeuille (2020) for more details). None of the treated states has a negative weight, suggesting that contamination due to the staggered adoption of mandated coverage is not a threat to our estimates.

[33]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 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, including state-specific intercepts interacting with time-varying coefficients (Bai, 2009). GSC links the matching and interactive fixed-effect methods and brings together synthetic control and interactive fixed-effect models, where the DD model is a particular case. For a review of recent studies on synthetic control methods, see Abadie (2020).

factuals. 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 year 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.[34] The vector X_{it} is a set of time-varying state-level characteristics, including mothers' age, marital status, education, race, and fathers' race. We also include the state-level socioeconomic characteristics from the CPS data discussed above.

$\lambda'_i f_t$ denotes the interactive fixed effects where λ_i and f_t are r-vectors of state-specific intercepts and time-varying coefficients, respectively, 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 determine which model fits better.[35] Details of the estimation procedure of our GSC framework are provided in 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 the 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.[36] We aggregate the data into state-year cells and estimate the model separately for each generosity level

[34] As noted above, we follow Schmidt (2007) and allow the mandated coverage to affect the incidence of multiple births with a two-year delay.

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

[36] 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.

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 inferences 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.[37]

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 of 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 the mandate dates for states grouped together are generally very close to each other (see Table 1). However, we also estimate the GSC model separately for each state, comparing it to never mandated states, and present results in Figure D.1 in Appendix D.

4.4 GSC Results

Plots presenting the estimated counterfactual and estimated effects on the treated states for each level of coverage are presented in Figure 4 and suggest that the GSC estimator 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.[38] 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, for comparison with the previous literature on infertility mandates. The first column shows that any mandated coverage increases the multiple birth rate by 0.10 percentage

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

[38] See Figure C.1 in Appendix C for a graphical representation of the estimates.

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 intensive infertility treatment only (level 0, with no coverage for IVF) does not affect the multiple birth rate relative to states that never enact mandates. This finding is relatively consistent across our results. Panels C through G show that states with more generous coverage generally 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 0.24 percentage point increase (23.07%) in states with level 5 coverage.[39]

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 5 and Figure 6. After controlling for covariates, the estimated effects for women 35 and older tend to be larger than those for 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 0.59 percentage point (46.45%) increase in states with level 5 coverage.[40] The estimated effects for younger women, especially for high generosity states, are much smaller and less statistically significant (for example, a 0.11 percentage point (10.78%) increase in level 5 states). The estimated effects from level 2 coverage, which includes only Connecticut and limits the number of transferred embryos to two, are relatively small and insignificant. This could imply that limiting the number of transferred embryos could be an effective policy intervention for decreasing the incidence of multiple births.

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

[39] The point estimate for level 1 coverage is negative in Column 1 without covariates but becomes positive in Column 2 when covariates are added

[40] The point estimate of level 1 coverage on older women is negative but not statistically different from zero. Since these states only cover one cycle, this could be due to older women needing more than one cycle and/or to fewer older women using the treatment.

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 consistent with those from the multiple birth rate analysis. 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 2.71 infants (25.69%) in states with level 5 coverage, and again, the effects are larger for older women.

Table C.1 presents p-values for the two-sample Welch statistics testing $H_0 : \delta_{Level_i} = \delta_{Level_{i+1}}$ versus $H_1 : \delta_{Level_i} < \delta_{Level_{i+1}}$ for estimates with covariates presented in Table 4 and Table 5. i denotes the level of mandated IVF coverage. The tests assume that the population distributions are normal but have unequal variances. We can reject the null hypothesis with only a few exceptions, suggesting that in general estimated effects from higher levels of coverage are statistically larger, but the estimated effects are not strictly increasing.

As discussed above, we use the number of covered cycles as a proxy for generosity of the mandate, but there are other dimensions along which mandates vary. In Table D.1 in Appendix D, we examine whether the effect of any mandated coverage for IVF on multiple births varies depending on the presence of these other restrictions. These findings suggest that restrictions on waiting time, marital status, number of transferred embryos, and life time cap might affect the incidence of multiple births.

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.[41]

[41] 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). 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 the authors upon request.

4.5 Robustness analysis

We also estimate the effects of coverage generosity on the incidence of multiple births using a DD framework. In addition to serving as a 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. Our data go through 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.

We also estimate the effects of the generosity of mandated coverage on the incidence of multiple births using a DDD framework, further refining the treatment group by mothers' age (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 E.

Our estimates from the DD model are statistically significant and larger in magnitude than the estimates of Buckles (2013), who found small positive but insignificant effects on the overall multiple birth rate. This difference in findings could be driven by the states with the most recent and more generous mandated coverage, which were not included in the previous work. In Table E.2 in Appendix E, we show that when we limit the range of years to 1980-2002, the same years used by Buckles (2013), our estimated DD coefficients fall in size and are no longer statistically significant.

The overall story from the DD and DDD estimates confirms 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 to transfer more embryos per cycle. However, more generous IVF coverage could also incentivize new patients to seek treatment, and these new patients may have lower probabilities of success.[42]

We use two additional data sources to shed light on patients' behavior from mandated coverage generosity. First, we investigate patients' utilization behavior using fertility clinic-level data. 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 considered descriptive and do not provide causal estimates.

5.1 Evidence from IVF clinics

We use SART's clinic-level data from 1996 to 2010 to investigate the relationship between coverage generosity and patients' utilization behavior. Table A.1 presents summary statistics for these data. The average number of embryos transferred per cycle decreases over our study period in both mandated and never mandated states, likely due in part to changes in SART recommendations over time.[43] More embryos are transferred per cycle for women 35 and older than for younger women. In recent years, the share of cycles performed on women 35 and older is ten percentage points higher in the mandated states relative to the never mandated states.

Unfortunately, the mandate date for six out of the eight states with IVF coverage falls before the SART data is available. As a result, we cannot use a GSC model to examine

[42] Individuals on the margin of treatment could be those with lower probability of success, or those with lower ability to pay. Our descriptive results in the next section suggest the former, but we cannot fully tease out these two possibilities with our data.

[43] A major change to SART's guidelines occurred in 2004. We estimated an event study model and found 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. The estimates are available from the authors on request.

how coverage generosity affects patients' utilization behavior. 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.[44] This analysis should be considered purely descriptive and not intended to show causal effects. Instead, we view this analysis as complementary to our causal GSC estimates.

We exploit random variation between clinics within states in addition to variation across the states. An example of random variation between clinics might be doctors' opinions about the appropriate number of embryos to transfer, affecting the incidence of multiple births from an IVF cycle. We estimate a model specified as follows:

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

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 per clinic, 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 as defined above, 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 coefficients for all women, as well as broken out by

[44] 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).

age. These results suggest the following: First, more generous coverage is associated with a significantly higher number of cycles in a clinic, on average, which would lead to an increase in multiple births. Second, more generous coverage is associated with fewer transferred embryos per cycle for both older and younger women, with a stronger relationship for younger women. This, on its own, would imply a lower 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 a higher incidence of multiple births. Our GSC results using birth certificate data presented above show an overall causal increase in the incidence of multiple births, suggesting that the overall utilization effect, combined with the compositional effect, are likely to 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 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 might 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 fewer IVF cycles performed. Cohen and Chen (2010) find that mandated IVF coverage did not affect child adoption relative to never mandated states. However, the effects of mandated coverage on adoption could be 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 adopted children from birth to age six as such adoptions could be considered a substitute for conceiving through IVF in some circumstances.[45] Table A.2 in Appendix A presents descriptive statistics for these data. In the early years of our study period, the mandated states' adoption rate is higher than in the never mandated states. However, by the latter half of our time period, this pattern had reversed itself such 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 here because the mandate date for six out of the eight states that cover IVF 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 adopted children between birth and age six per ten thousand live births in each state-year cell. Table 7 presents the estimated coefficients, first for all women and then broken out by the women's age. 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 (from birth certificates, fertility clinics, and child adoptions) have three main takeaways. First, more generous IVF coverage increases the incidence of multiple births. Second, more generous coverage is associated with fewer transferred embryos for all women, but the association is stronger for younger women than older women. Third, more generous coverage is associated with both an increase in the number of cycles performed, as well as changes in the composition of patients, where the share of cycles performed on women over 35 years is greater in states with more generous coverage. This is mirrored by fewer child adoptions to older women

[45] We estimated our model for children older than six years, and the findings are similar. The estimates are available from the authors on request.

in states with more generous coverage. These findings provide suggestive evidence that differences in both overall utilization and in the composition of patients may dominate differences due to the transfer of fewer embryos per cycle, resulting in the overall increase in costly multiple births that we find in the birth certificate data.

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. We show that more generous coverage leads to higher rates of risky and costly multiple births. Our descriptive analysis using the SART data shows that while more generous coverage is associated with fewer embryos transferred per cycle, it is also associated with both more cycles per clinic and a larger share of cycles performed on older women. 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.[46]

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 intensity of treatments, for imposing a top-up price for more intense treatments, or for some combination of the two. In the IVF context, Hamilton et al.

[46] Our results are also consistent with the possibility that women in high generosity states are waiting longer to have children (e.g., Abramowitz (2017)).

(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. (2022) 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.

Abramowitz, J. (2020). Infertility insurance mandates and maternal mortality. *Consider This*.

Adashi, E. Y. and Gutman, R. (2018). Delayed Childbearing as a Growing, Previously Unrecognized Contributor to the National Plural Birth Excess. *Obstetrics & Gynecology*, 132(4):999–1006.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Barishansky, S. J., Hutchinson, A. P., Lawson, A. K., and Pavone, M. E. (2022). The parent trap: desire for multifetal gestation among patients treated for infertility. *Journal of Assisted Reproduction and Genetics*, pages 1–9.

Beemsterboer, S., Homburg, R., Gorter, N., Schats, R., Hompes, P., and Lambalk, C. (2006). The paradox of declining fertility but increasing twinning rates with advancing maternal age. *Human Reproduction*, 21(6):1531–1532.

Bhalotra, S., Clarke, D., and Mühlrad, H. (2022). Health and Labor Market Impacts of IVF Technology: Evidence from a Swedish Policy Mandate. *IZA Working Paper No. 14990*.

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.

de Chaisemartin, C. and D'Haultfoeuille, X. (2020). Two-way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9):2964–2996.

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. (2021a). Older yet fairer: how extended reproductive time horizons reshaped marriage patterns in israel. *American Economic Journal: Applied Economics*, 13(1):198–234.

Gershoni, N. and Low, C. (2021b). The power of time: The impact of free ivf on women's human capital investments. *European Economic Review*, 133:103645.

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. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

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.

Hazel, W. N., Black, R., Smock, R. C., Sear, R., and Tomkins, J. L. (2020). An age-dependent ovulatory strategy explains the evolution of dizygotic twinning in humans. *Nature Ecology & Evolution*, 4(7):987–992.

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.

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.

Roth, J., Sant'Anna, P. H. C., Bilinski, A., and Poe, J. (2022). What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature.

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

	Mandate year	Coverage level	Number of covered cycles	Long min infertility time	Religious exemption	Firm size exemption	Marital status restriction	Age restriction	Restricted embryo transfers	Lifetime spending cap
Montana	1987	0	0	-	-	-	-	-	-	-
New York	1990	0	0	-	-	-	-	-	-	-
Ohio	1991	0	0	-	-	-	-	-	-	-
West Virginia	1995	0	0	-	-	-	-	-	-	-
Arkansas	1987	1	-	X (2 years)	-	-	X	-	-	$15,000
Hawaii	1989	1	1	X (5 years)	-	-	X	-	-	
Connecticut	2005	2	2	-	X	-	-	X ($\leq$ 40)	X (2 embryos)	
Maryland	1985	3	3	-	X	X ($\leq$ 50)	-	-	-	$100,000
Rhode Island	1989	3	-	-	-	-	-	X (25-40 years)	-	$100,000 + 20% co-payment
Illinois	1991	4	4	-	X	-	-	-	-	-
New Jersey	2001	4	4	X (2 years)	X	X ($\leq$ 50)	-	X ($\leq$ 46 years)		-
Massachusetts	1987	5	Unlimited	-	-	X ($\leq$ 25)	-	-	-	-

Note: "X" denotes states with a specific aspect of mandated IVF coverage. States with "Long minimum infertility time" are those which require more than one year of infertility to be eligible for coverage. States with "Religious exemption" are those which do not require religious organizations to provide coverage. In states with "Firm size exemption," employers with fewer than a certain number of employees do not have to provide coverage. States with "Age restrictions" impose an age restriction for coverage eligibility. Connecticut restricts the number of embryos transferred.
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 the likelihood of a multiple birth, 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)
Age cutoff	40	40	40	40	46	46	40	40
Coverage level	2	2	3	3	4	4	0	0
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
Covariates included	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 the likelihood of a multiple birth 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 use data on all 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 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	900
	(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	0.42***	0.24***	0.87***	0.59**	0.27**	0.11	1,080
	(0.09)	(0.08)	(0.35)	(0.30)	(0.13)	(0.08)	
Pre-mandate mean	1.04	1.04	1.27	1.27	1.02	1.02	
	(0.08)	(0.08)	(0.30)	(0.14)	(0.08)	(0.08)	
Covariates included	No	Yes	No	Yes	No	Yes	
State and time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	

Notes: This table presents the estimated average treatment effect on the treated from the GSC model specified in Equation (2). Data are 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. See Figure C.1 in Appendix C for graphical presentation of this table.

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

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 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	900
	(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	4.60***	2.71***	10.04**	9.72***	2.31	1.64**	1,080
	(1.04)	(0.90)	(3.23)	(3.84)	(1.57)	(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)	(3.41)	(0.79)	(0.82)	
Covariates included	No	Yes	No	Yes	No	Yes	
State and time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	

Note: See notes for Table 4.

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

Table 6: Association between IVF coverage generosity level and patients' IVF utilization behaviour, Mixed Effects model

	All women				Women 35 and older				Women under 35	
	Total number of cycles per clinic		Average # of transferred embryos per cycle		Share of cycles to women 35+		Average # of transferred embryos per cycle		Average # of transferred embryos per cycle	
	(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
Covariates included	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
P-value of $Ch2$ stat		0.00		0.00		0.00		0.00		0.00
Number of observations	4576	4,576	3,821	3,821	4,574	4,574	3,822	3,822	4,562	4,562

Notes: This table presents the estimated association between the generosity of IVF coverage and patients' utilization behavior using IVF clinic data using the ME model specified in Equation (3). 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. The $Ch2$ statistic is used to test the null hypothesis that the estimated coefficients are all equal. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table 7: Association between IVF coverage generosity level and adopted children per ten thousand live births, Mixed Effects model

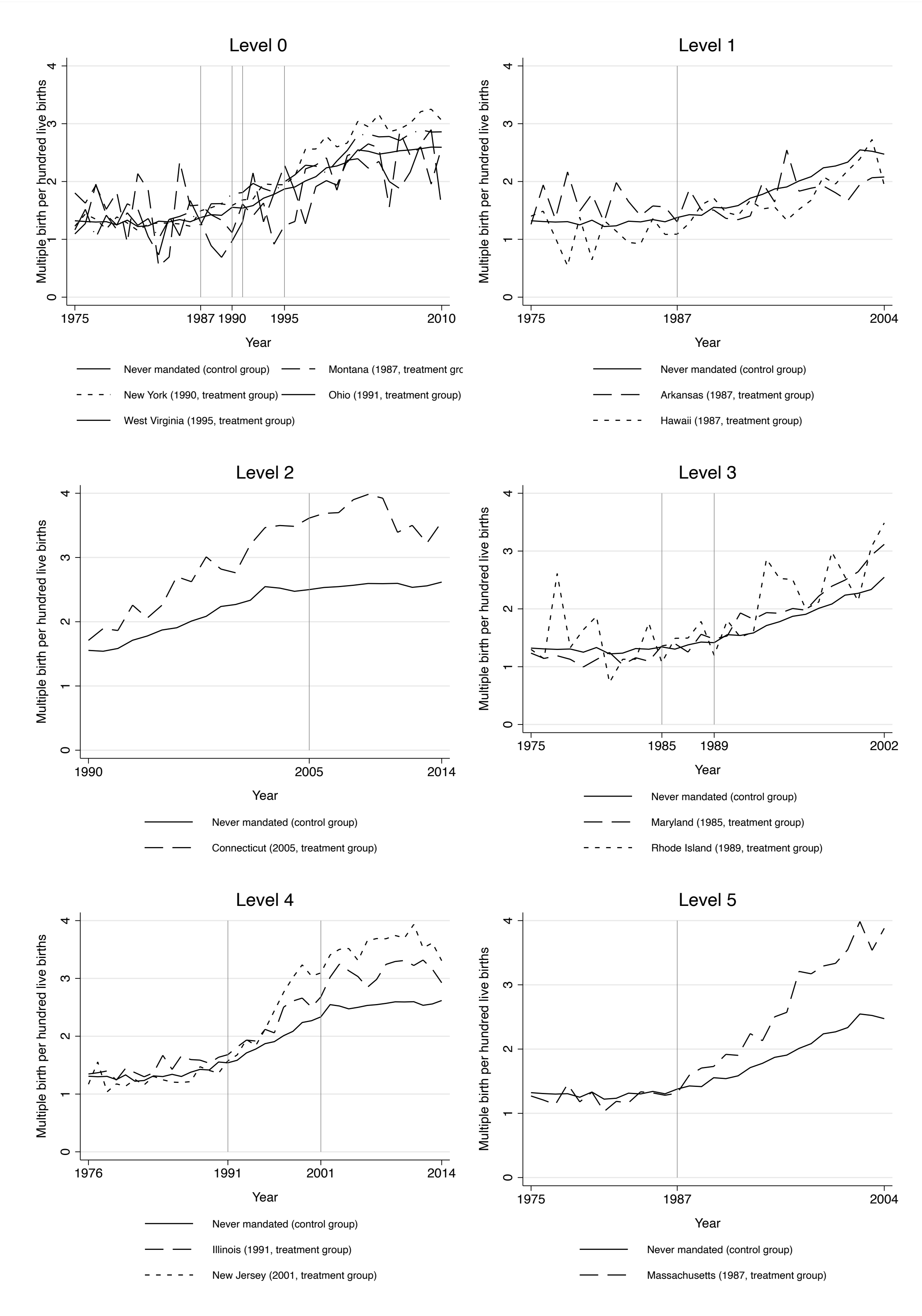
	All women		Women 35 and older		Women under 35	
	(1)	(2)	(3)	(4)	(5)	(6)
All levels	3.11		-11.83		-0.04	
	(6.79)		(52.77)		(1.23)	
Level 0		9.72		170.11		0.17
		(14.25)		(136.52)		(3.11)
Level 1		-29.76		-143.18		-5.36
		(19.23)		(208.76)		(5.66)
Level 2		16.57***		76.91**		1.33
		(6.26)		(39.23)		(1.65)
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
Covariates included	Yes	Yes	Yes	Yes	Yes	Yes
P-value of $Ch2$ stat		0.00		0.00		0.52
Observations	906	906	906	906	883	883

Note: This table presents the estimated association between the generosity of IVF coverage and child adoption using adoption data and the ME model specified in Equation (3). The data include children ages 0-6 adopted between 1994 to 2014. 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. The $Ch2$ statistic is used to test the null hypothesis that the estimated coefficients are all equal. $*p < 0.10, **p < 0.05, ***p < 0.01$

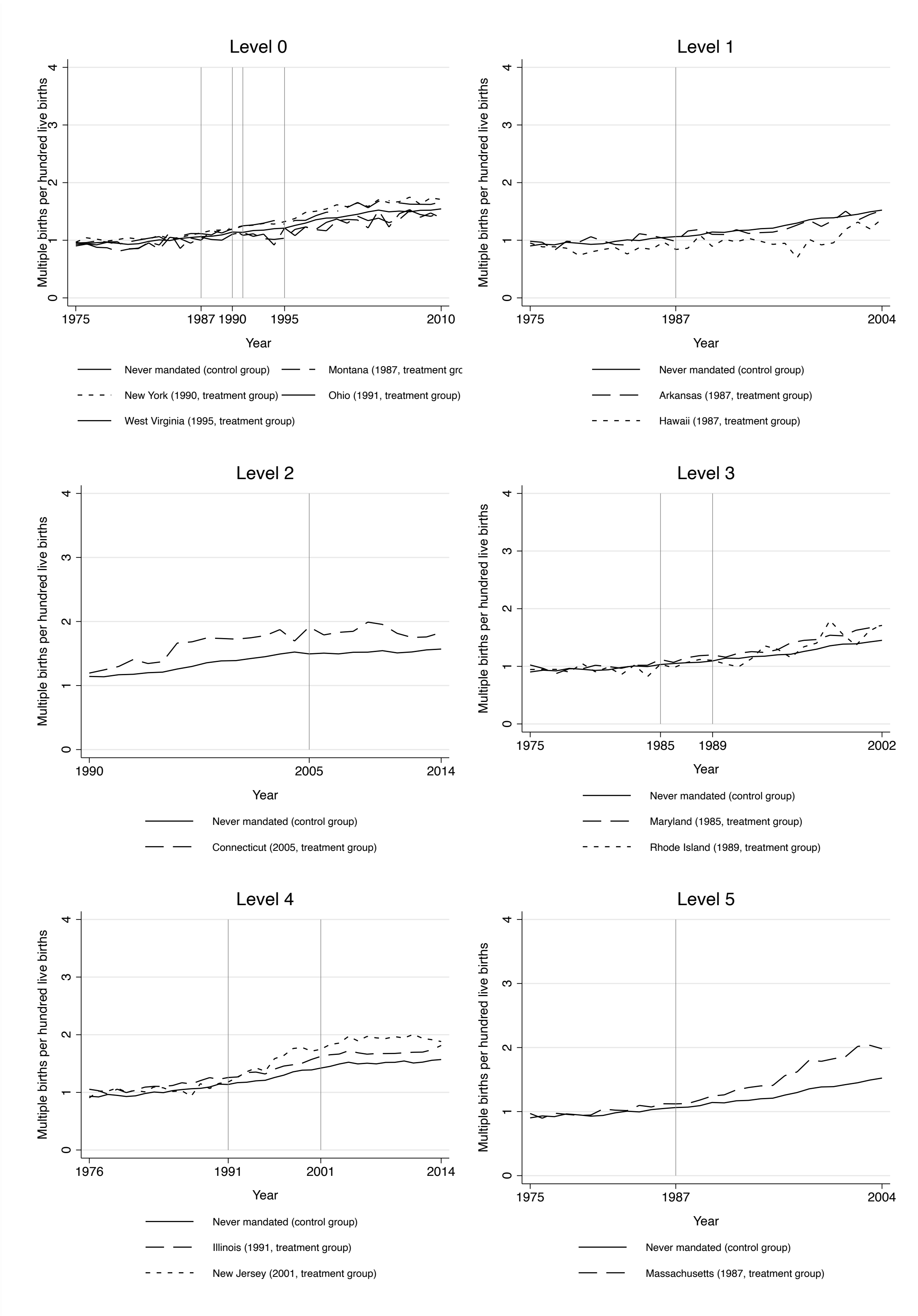
Figures

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

(a) Women 35 and older

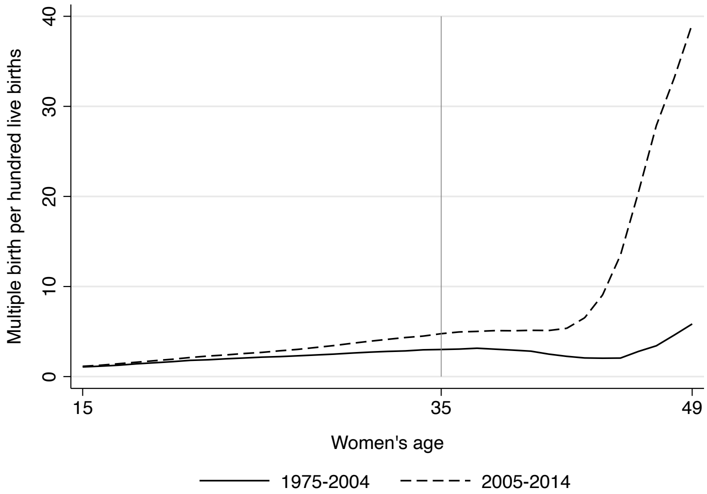


(b) Women under 35



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.

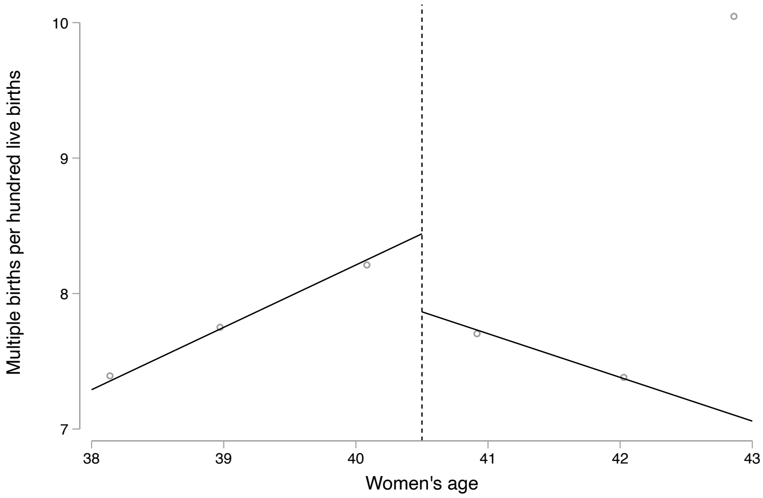
Figure 2: Multiple births per hundred live births by women's age



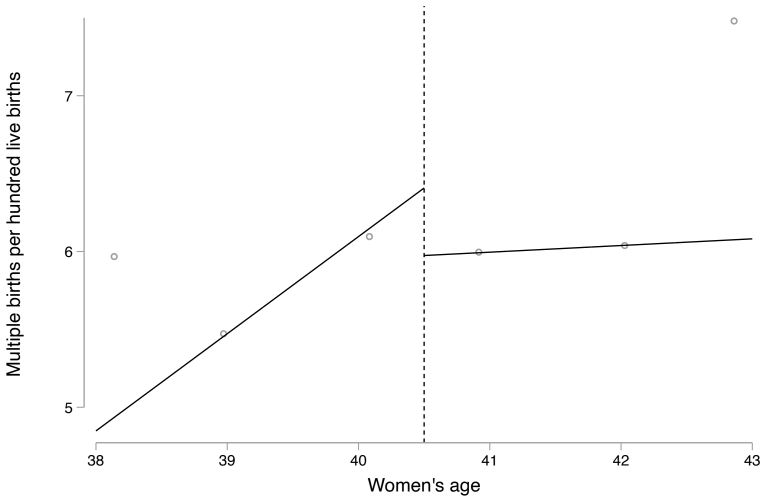
Note: Authors' calculations from the Detail Natality data. Multiple births are defined as births that are not singleton.

Figure 3: RDD plots of multiple births per hundred live births by women's age

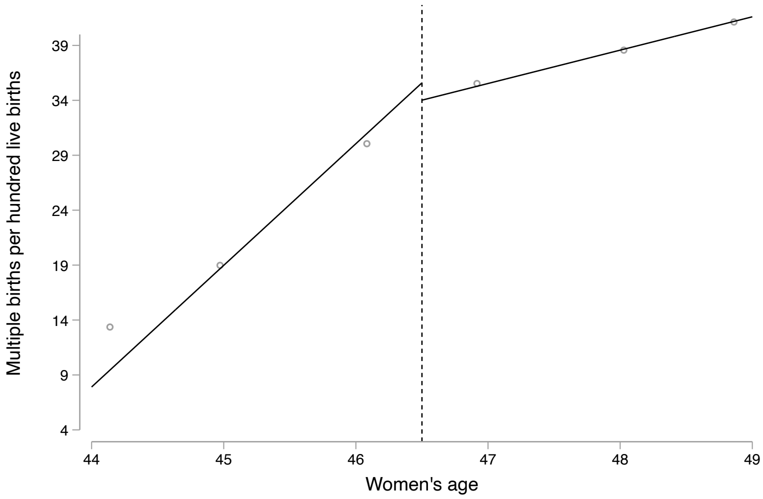
(a) Connecticut



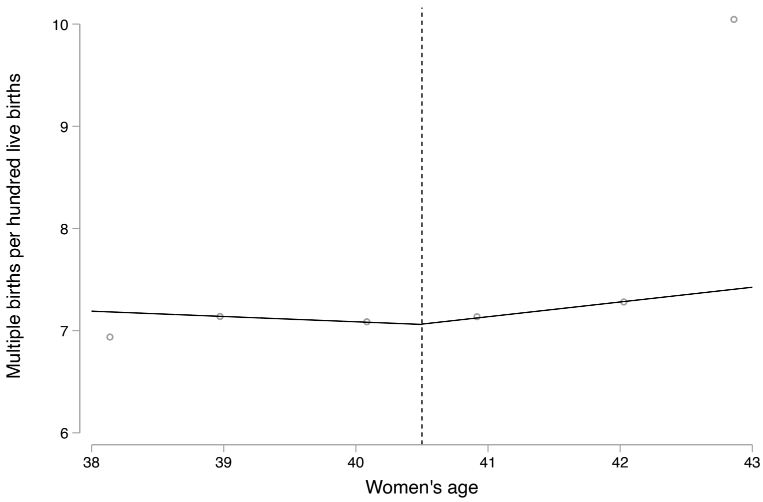
(b) Rhode Island



(c) New Jersey



(d) New Jersey (placebo)

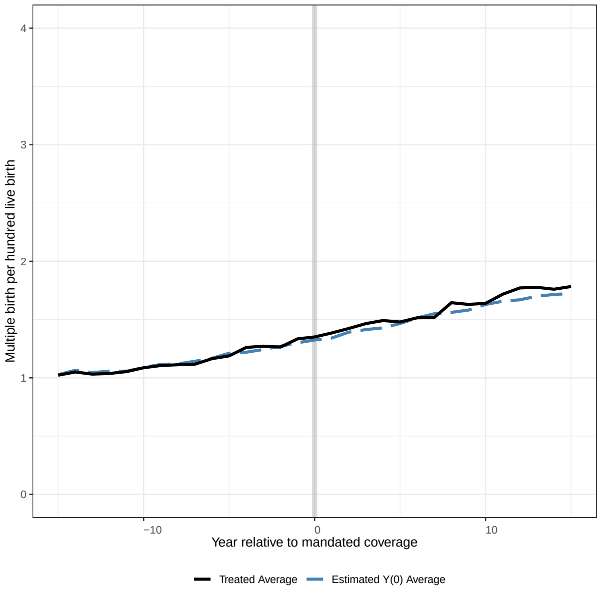


Note: This figure plots the multiple births per hundred live births by women's age two years after mandated IVF coverage in Connecticut (2007–2014), Rhode Island (1991–2014), and New Jersey (2003–2014).

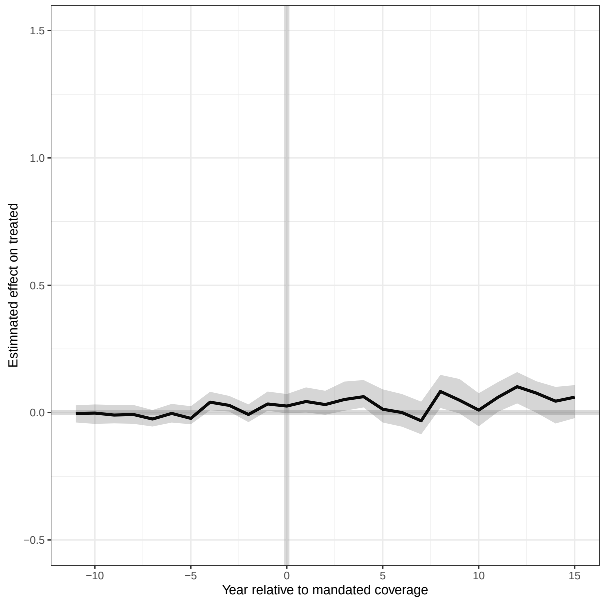
Figure 4: Effects of IVF coverage generosity level on multiple births per hundred live births, GSC model for all women

(a) All levels

(1) Treated average and estimated average for treated states

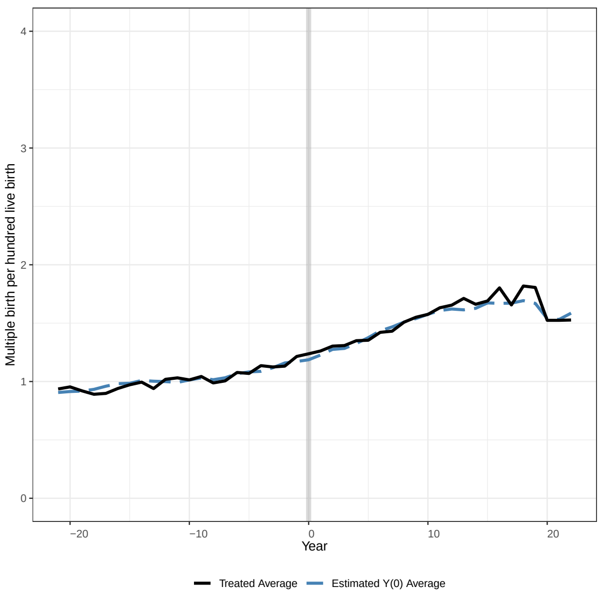


(2) Estimated treatment effect on treated

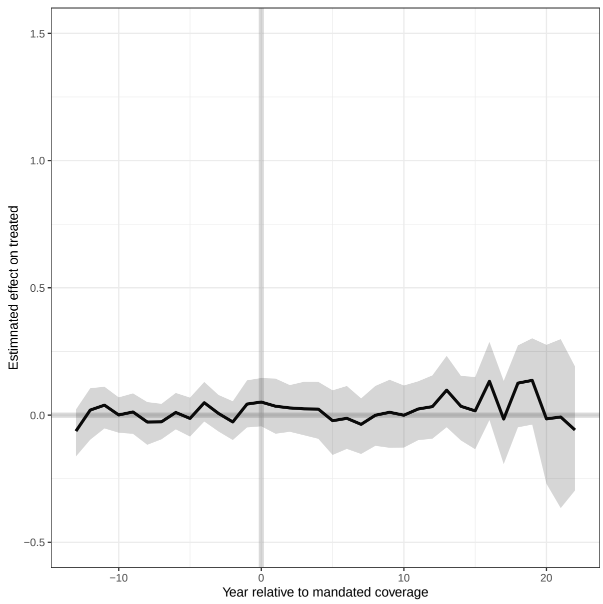


(b) Level 0

(1) Treated average and estimated average for treated states

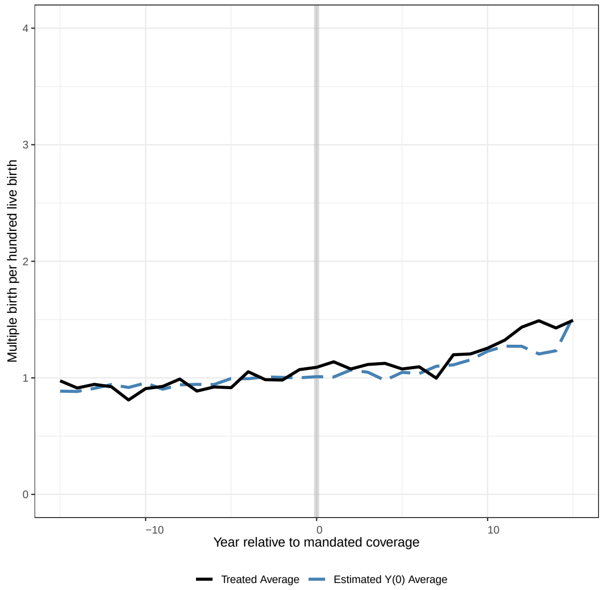


(2) Estimated treatment effect on treated

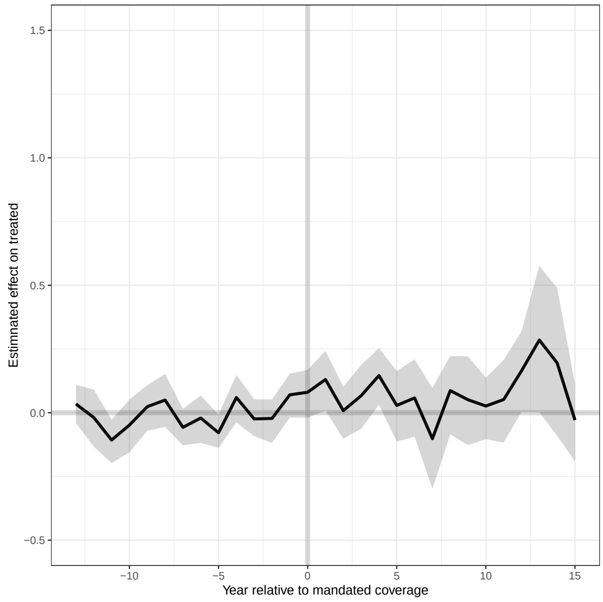


(c) Level 1

(1) Treated average and estimated average for treated states

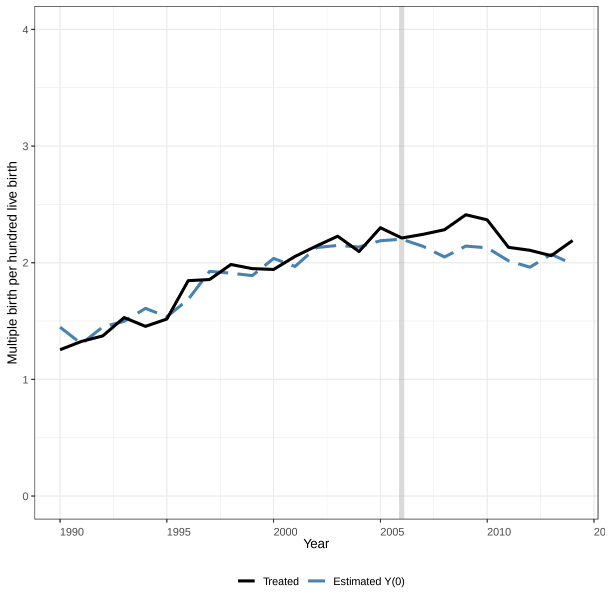


(2) Estimated treatment effect on treated

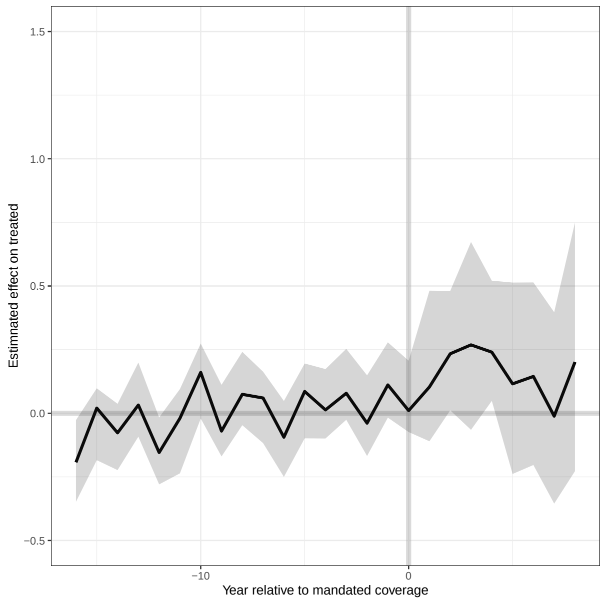


(d) Level 2

(1) Treated average and estimated average for treated states

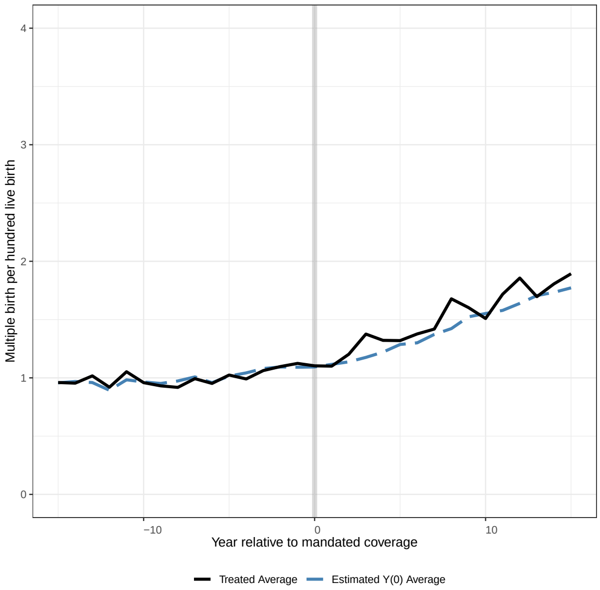


(2) Estimated treatment effect on treated

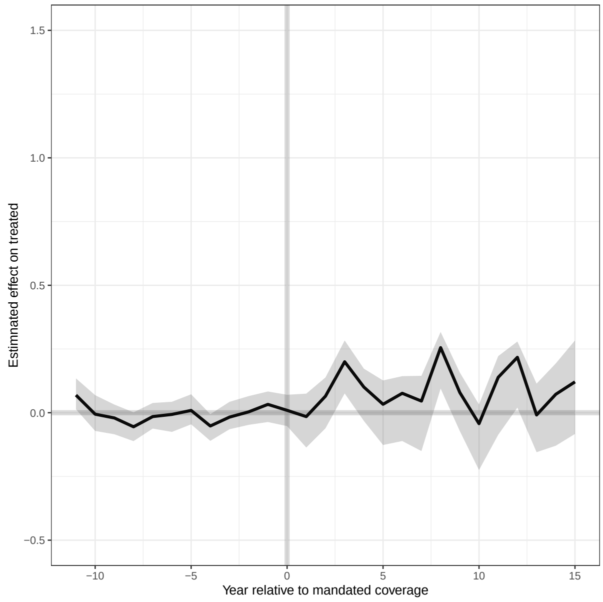


(e) Level 3

(1) Treated average and estimated average for treated states

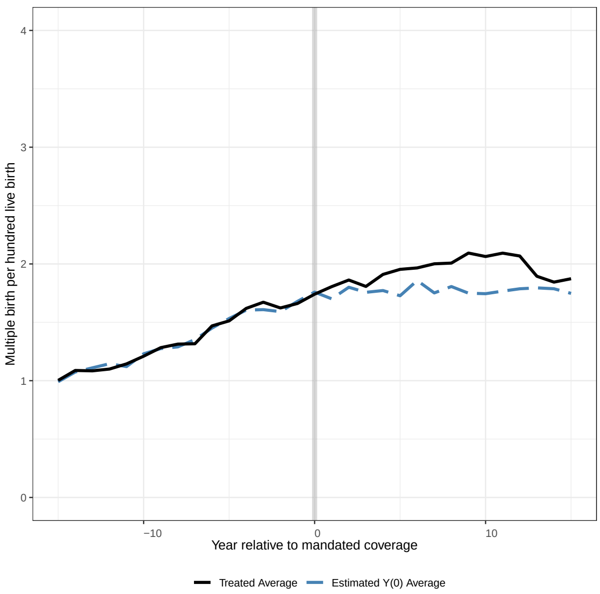


(2) Estimated treatment effect on treated

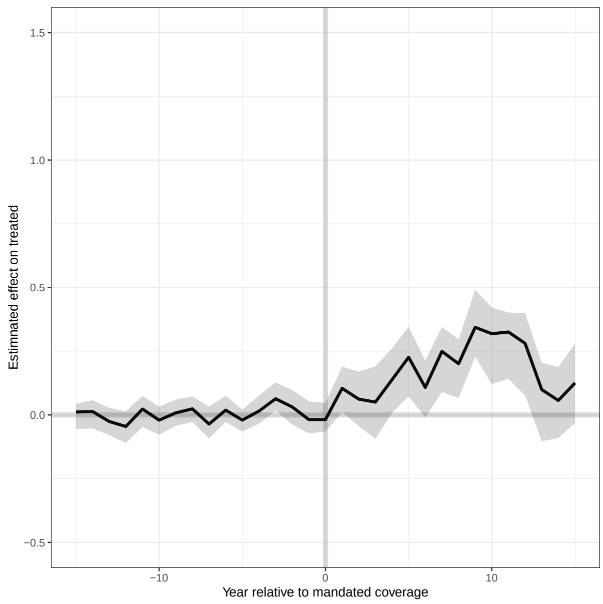


(f) Level 4

(1) Treated average and estimated average for treated states



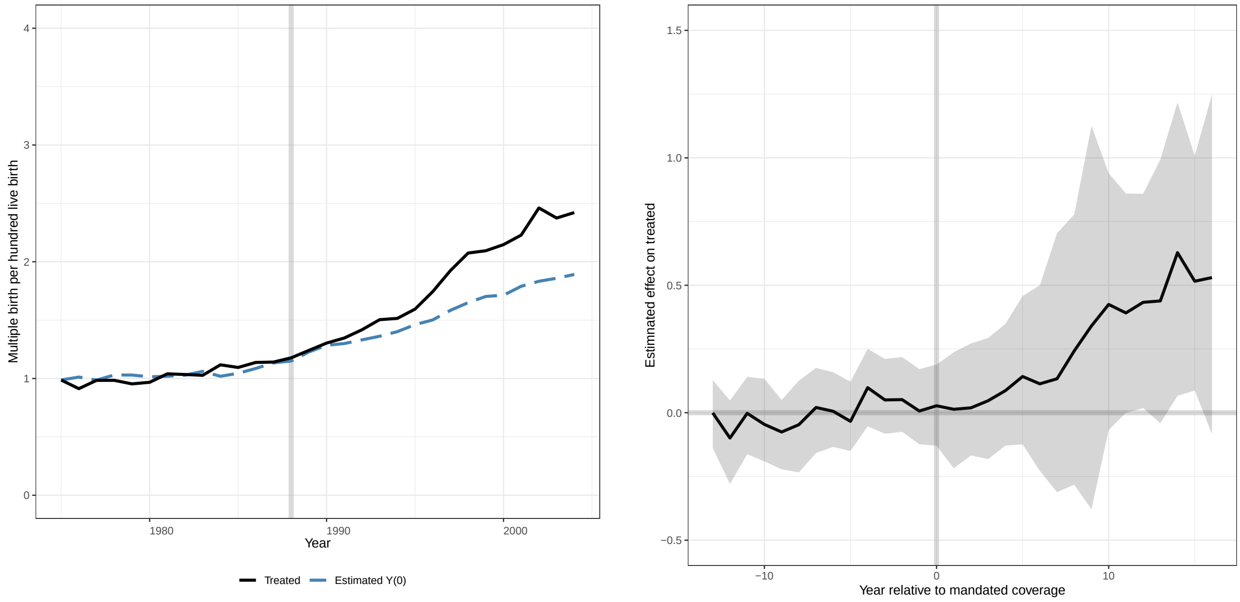
(2) Estimated treatment effect on treated



(g) Level 5

(1) Treated average and estimated average for treated states

(2) Estimated treatment effect on treated

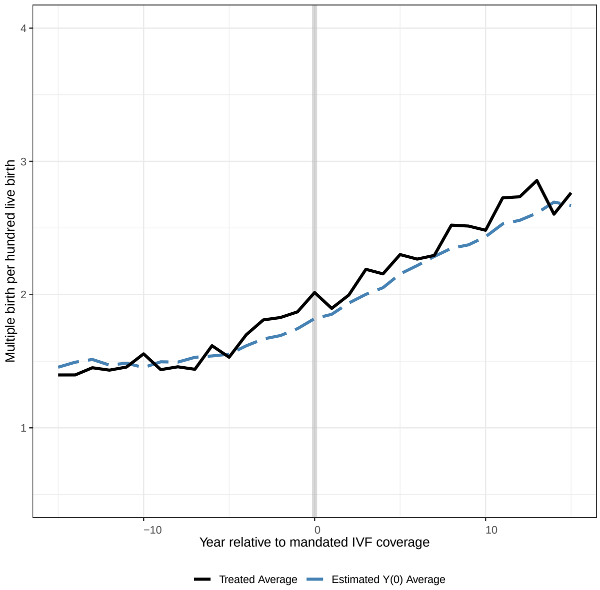


Note: This figure plots the estimated counter-factual outcome $Y(0)$ and the treatment effect on the treated for multiple births per hundred live births using the GSC model specified in Equation (2). The sample includes all births in the US from 1975-2014 from the National Vital Statistics, aggregated 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.

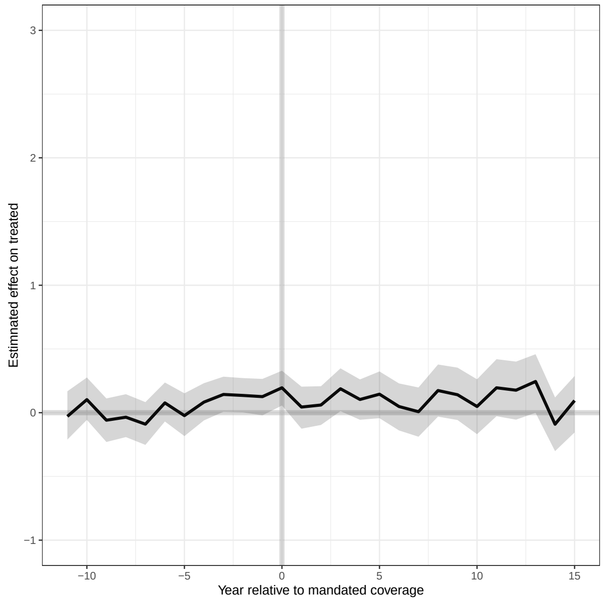
Figure 5: Effects of IVF coverage generosity level on multiple births per hundred live births, GSC model, women 35 and older

(a) All levels

(1) Treated average and estimated average for treated states

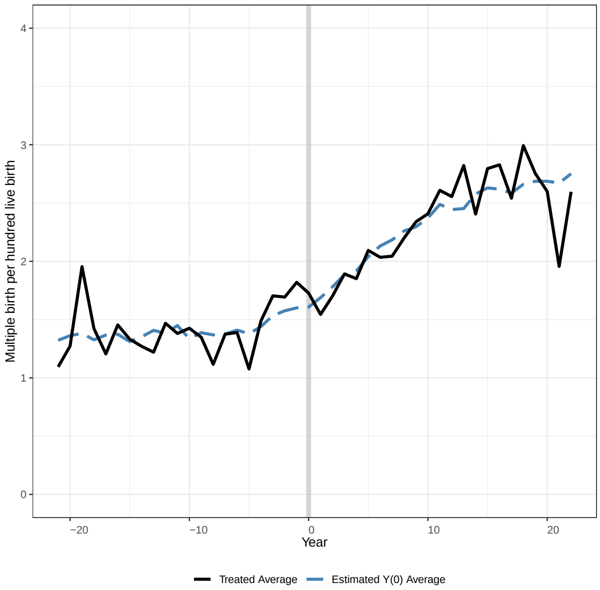


(2) Estimated treatment effect on treated

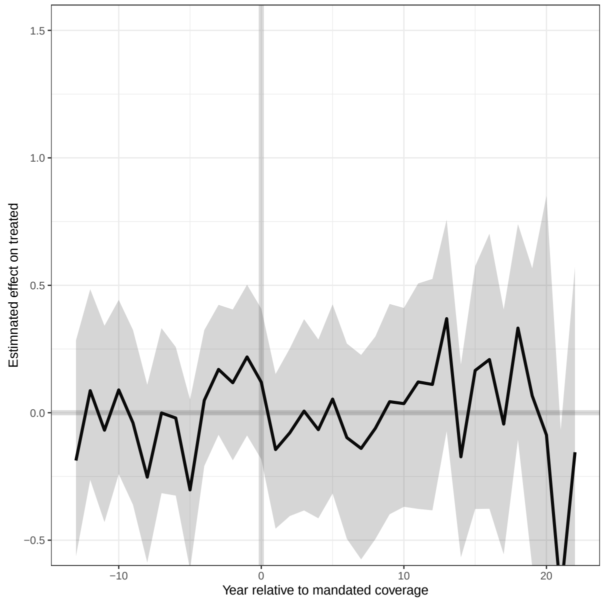


(b) Level 0

(1) Treated average and estimated average for treated states

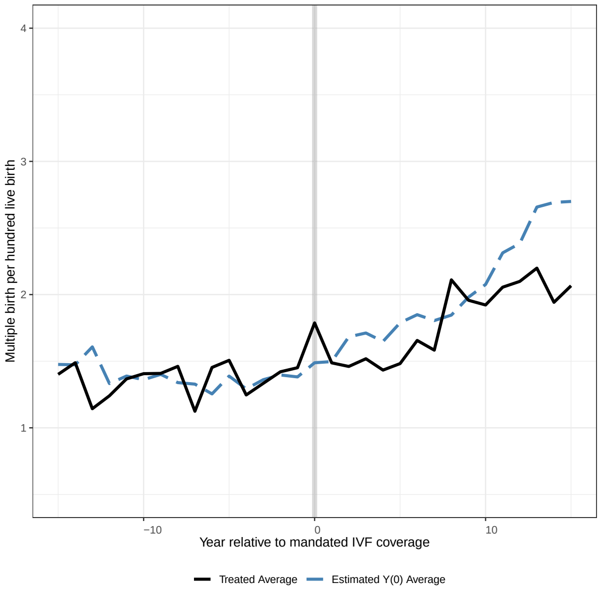


(2) Estimated treatment effect on treated

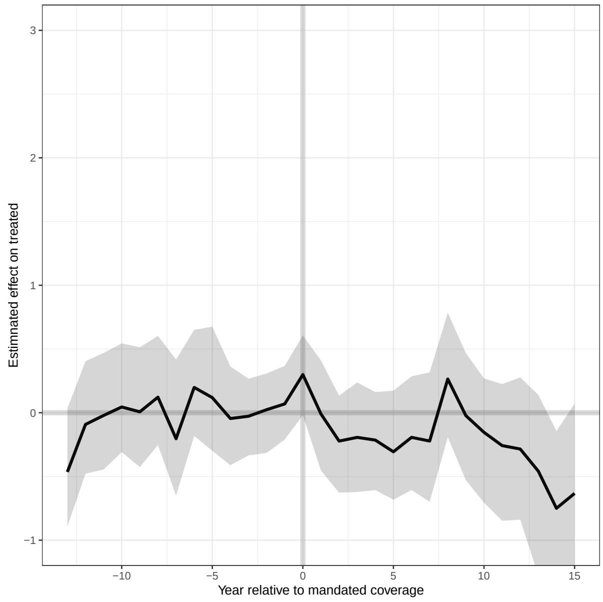


(c) Level 1

(1) Treated average and estimated average for treated states

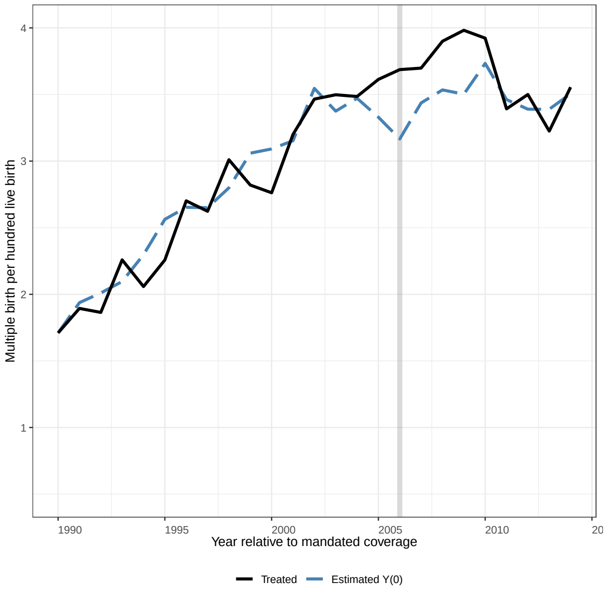


(2) Estimated treatment effect on treated

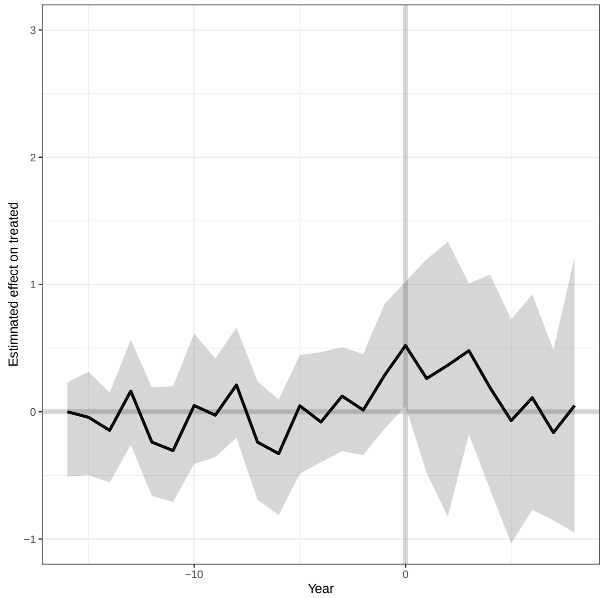


(d) Level 2

(1) Treated average and estimated average for treated states

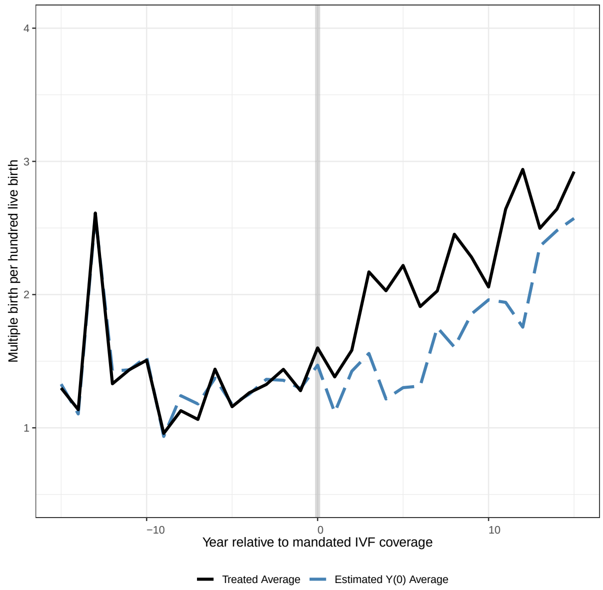


(2) Estimated treatment effect on treated

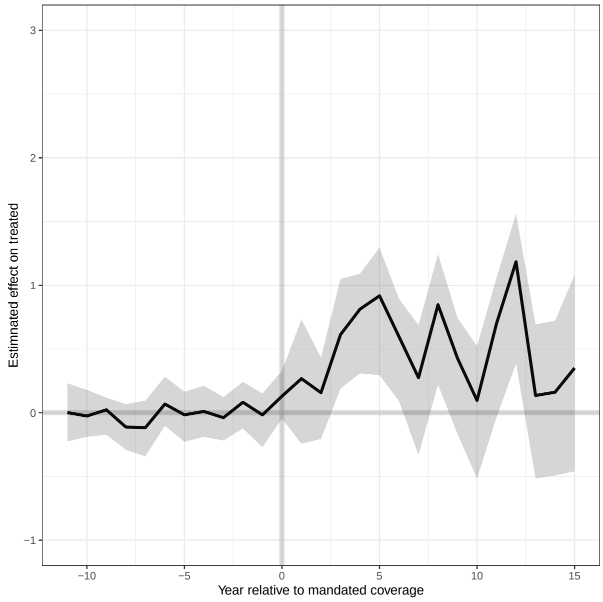


(e) Level 3

(1) Treated average and estimated average for treated states

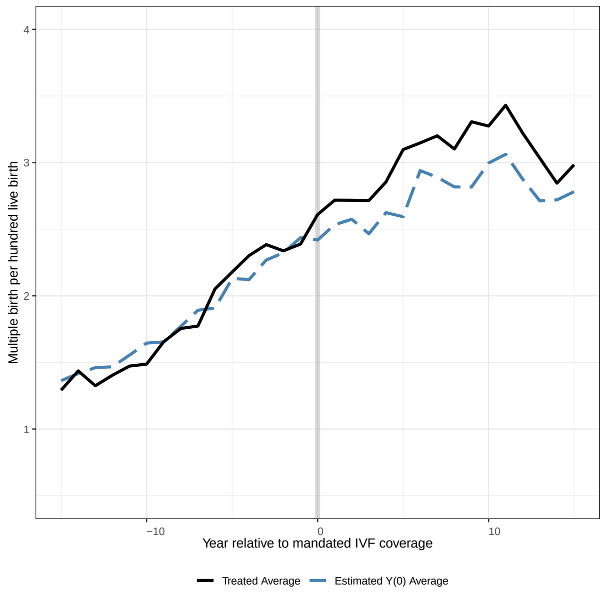


(2) Estimated treatment effect on treated

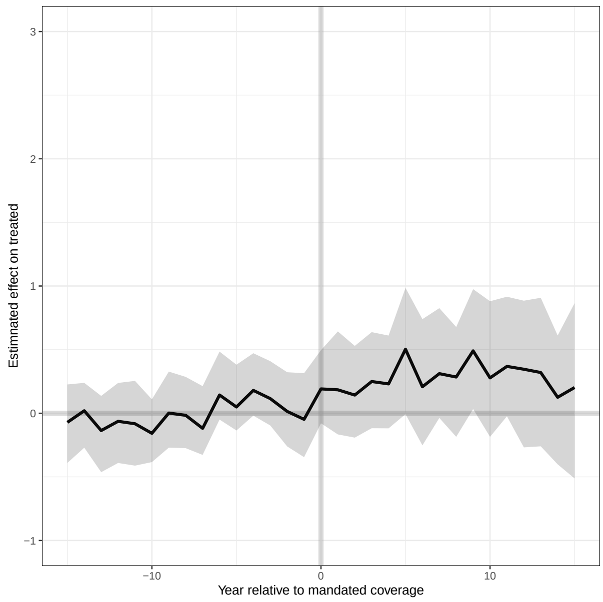


(f) Level 4

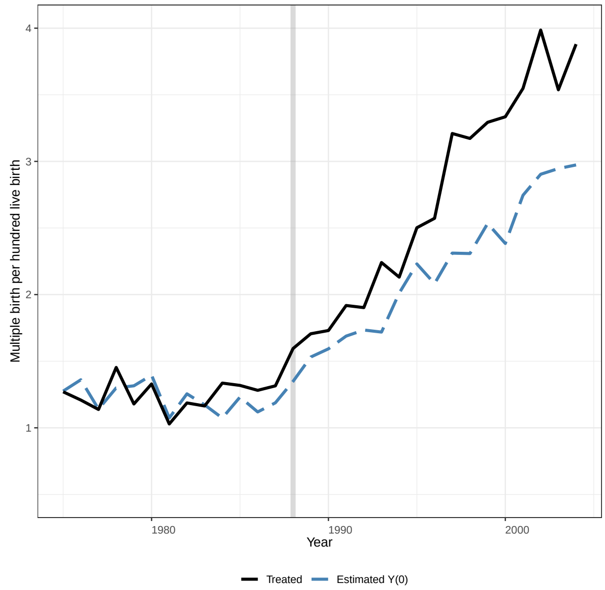
(1) Treated average and estimated average for treated states

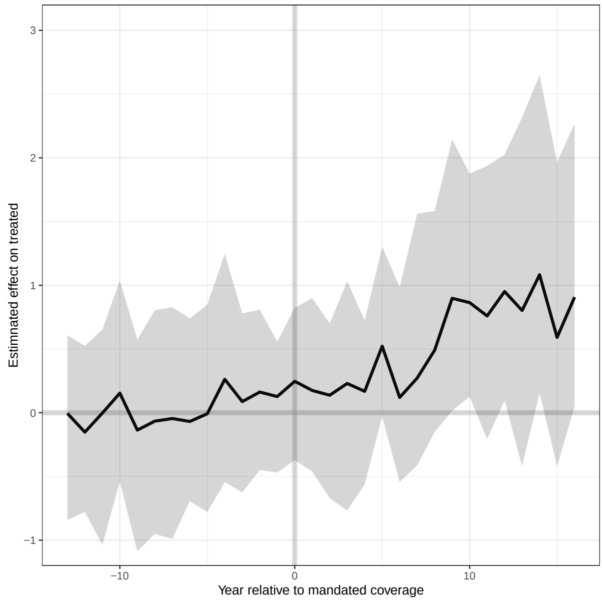


(2) Estimated treatment effect on treated



(g) Level 5



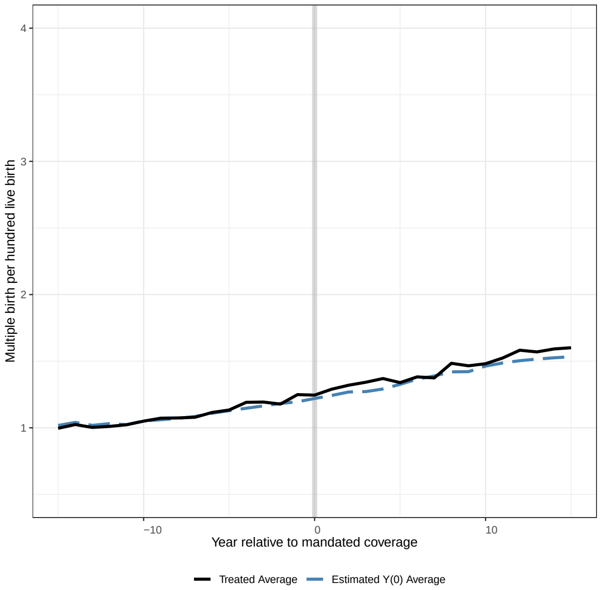


Notes: See notes for Figure 4.

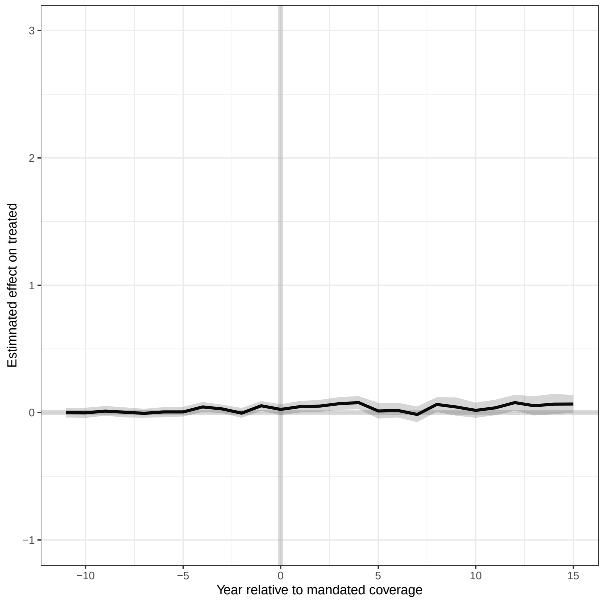
Figure 6: Effects of IVF coverage generosity on multiple births per hundred live births, GSC model, women under 35

(a) All levels

(1) Treated average and estimated average for treated states

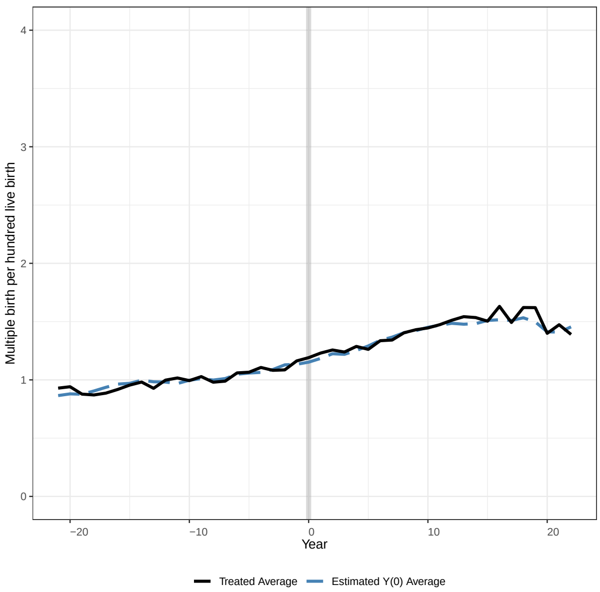


(2) Estimated treatment effect on treated

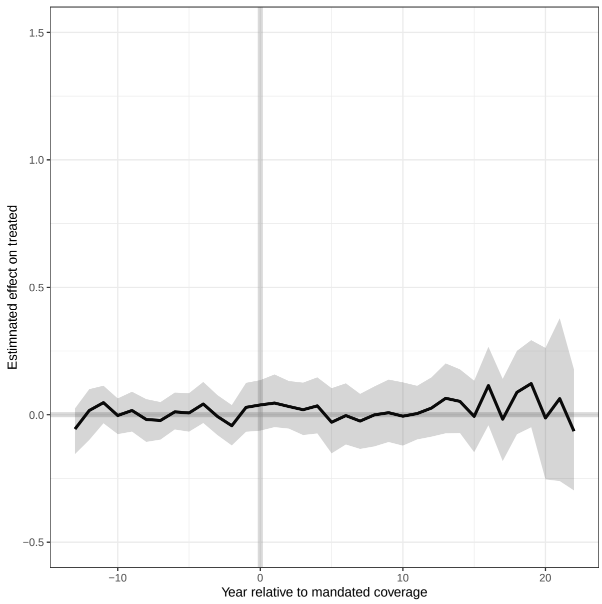


(b) Level 0

(1) Treated average and estimated average for treated states

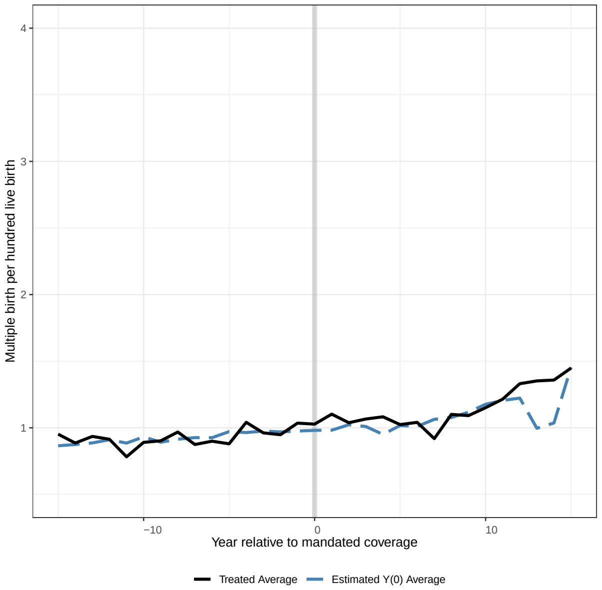


(2) Estimated treatment effect on treated

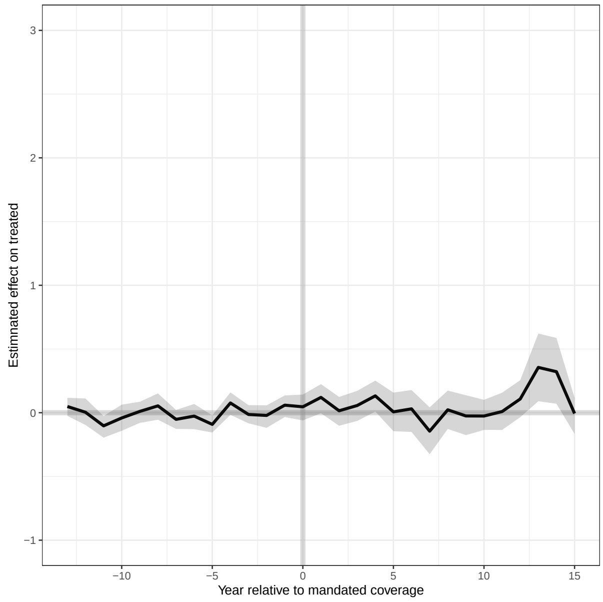


(c) Level 1

(1) Treated average and estimated average for treated states

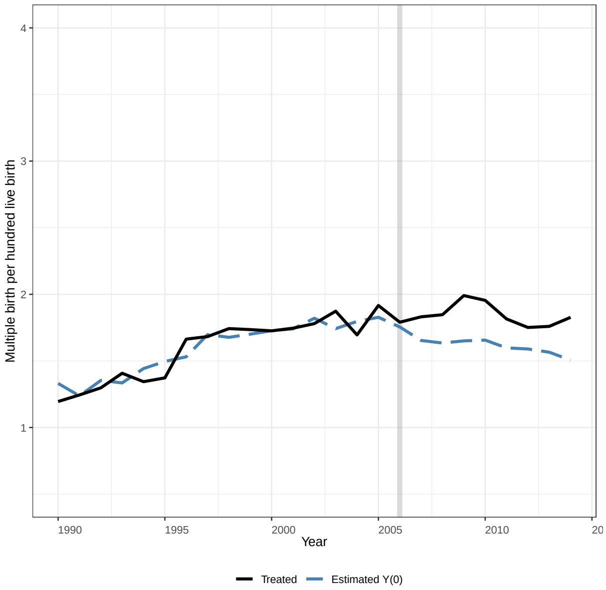


(2) Estimated treatment effect on treated

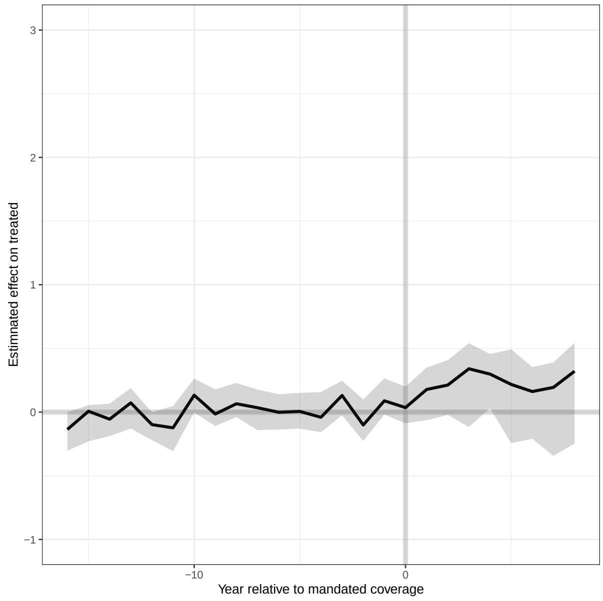


(d) Level 2

(1) Treated average and estimated average for treated states

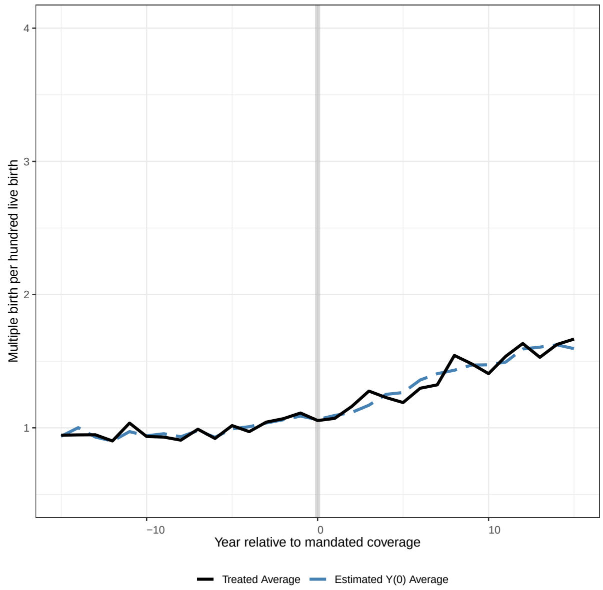


(2) Estimated treatment effect on treated

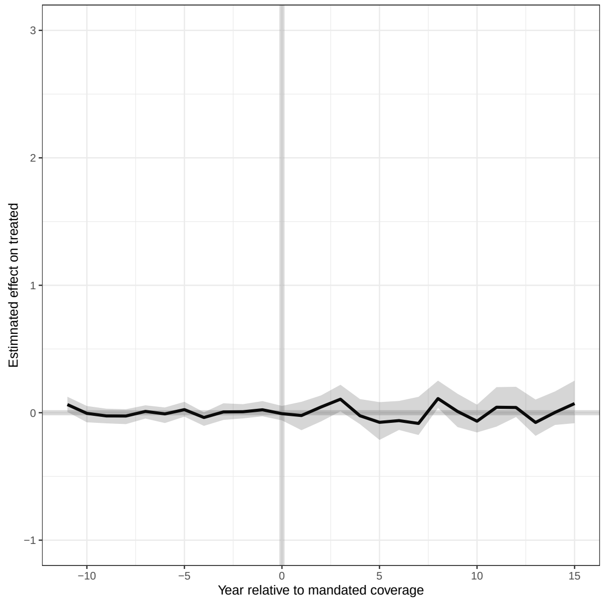


(e) Level 3

(1) Treated average and estimated average for treated states

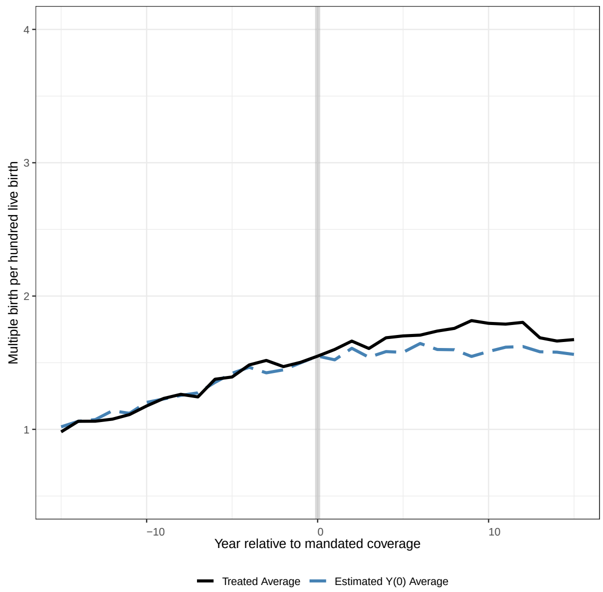


(2) Estimated treatment effect on treated

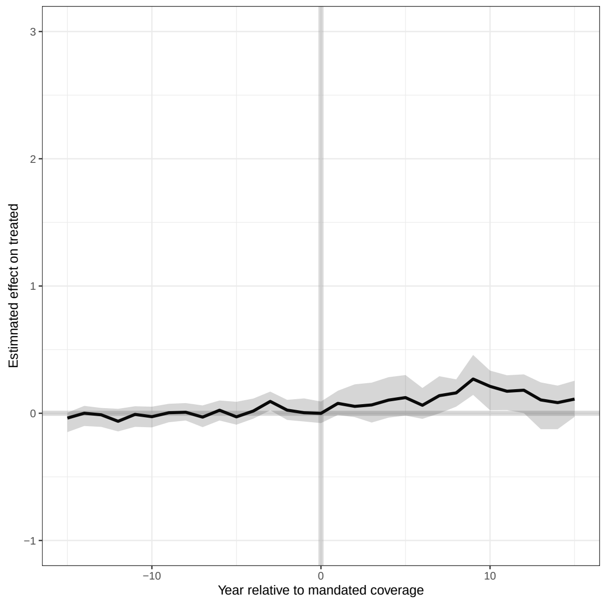


(f) Level 4

(1) Treated average and estimated average for treated states

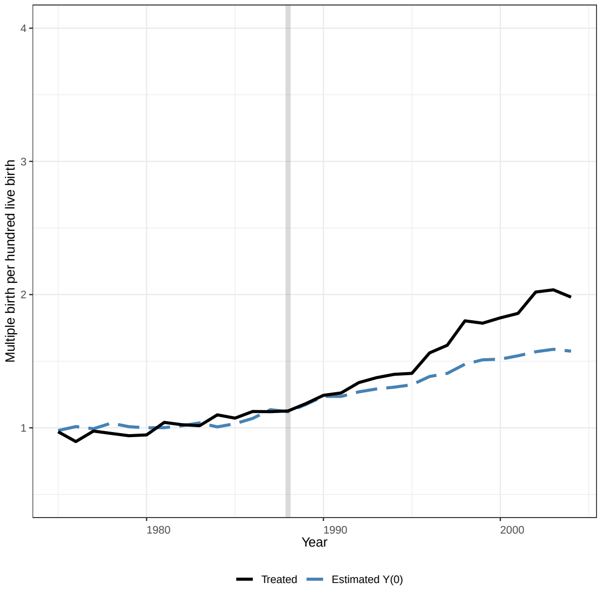


(2) Estimated treatment effect on treated

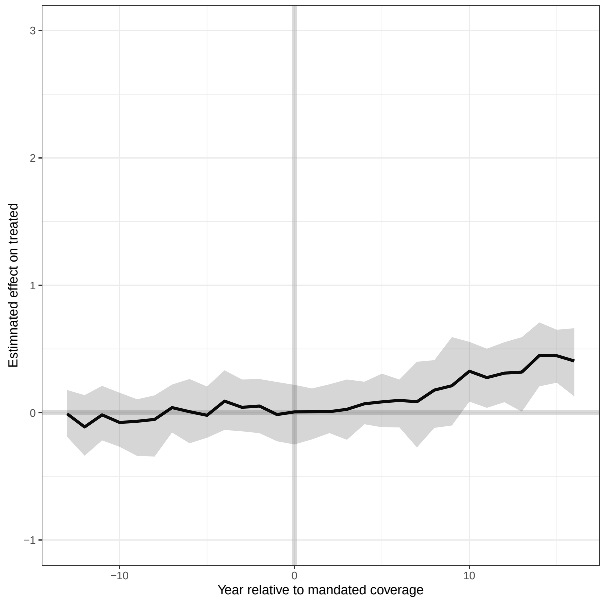


(g) Level 5

(1) Treated average and estimated average for treated states



(2) Estimated treatment effect on treated



Notes: See notes to Figure 4.

Appendix

A Summary statistics

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

	Never mandated states (control group)		*Mandate to cover states (treatment group)*	
	1996-2004	2005-2010	1996-2004	2005-2010
Total number of cycles	231,699	201,147	199,085	176,306
Average number of embryos transferred for all women	3.25	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 Annual Social and Economic Supplement (CPS)

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

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

B Estimation procedure of a GSC model

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

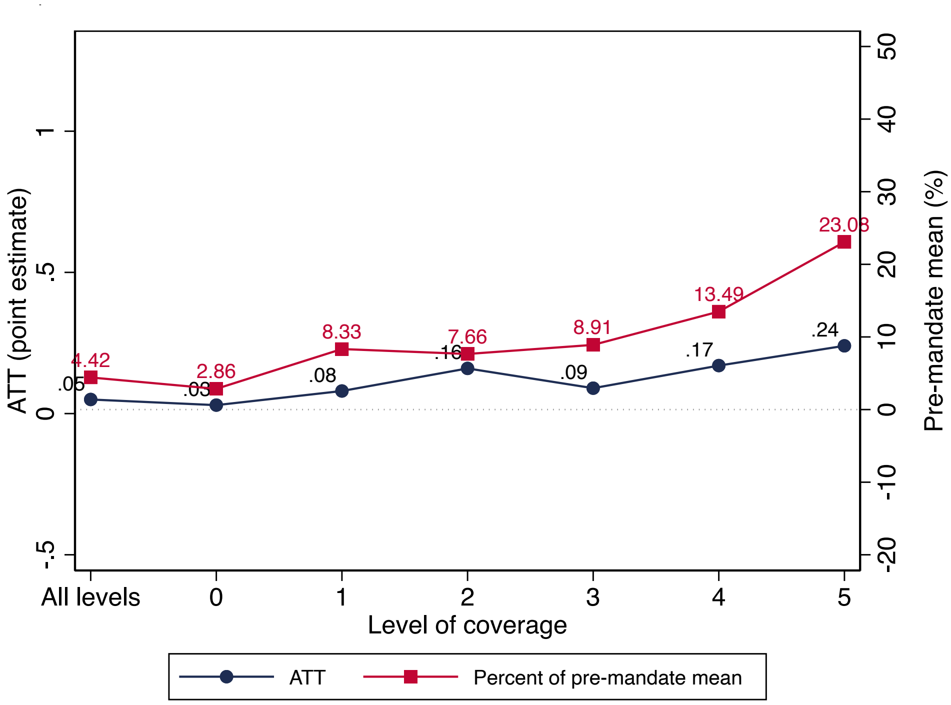
The estimated Average Treatment effect on Treated at time t, ATT_t then is:

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

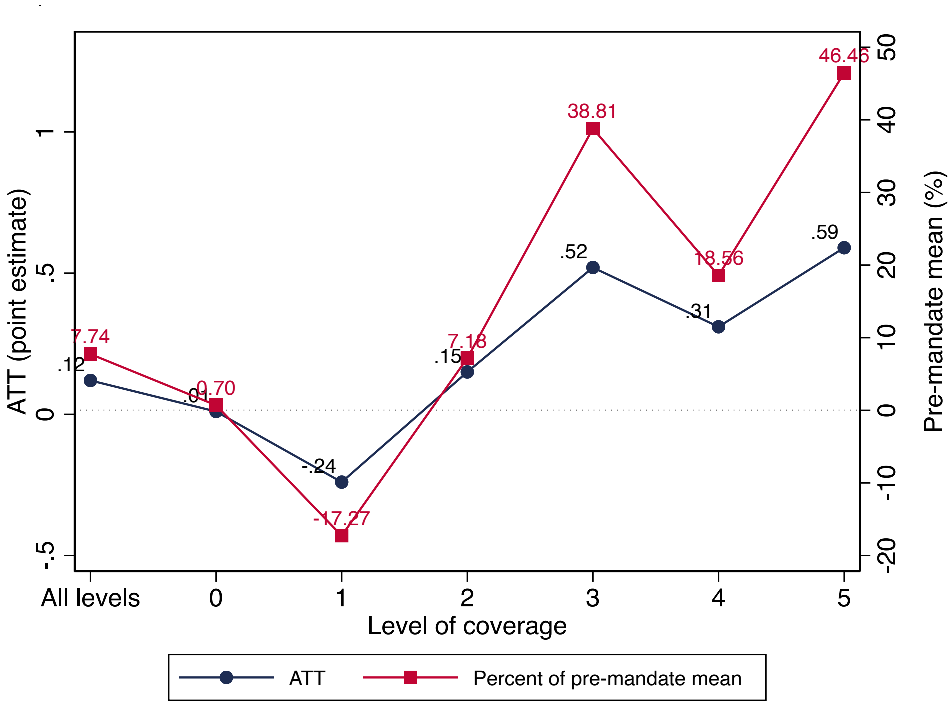
C Graphical presentation and statistical significance tests of GSC estimates

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

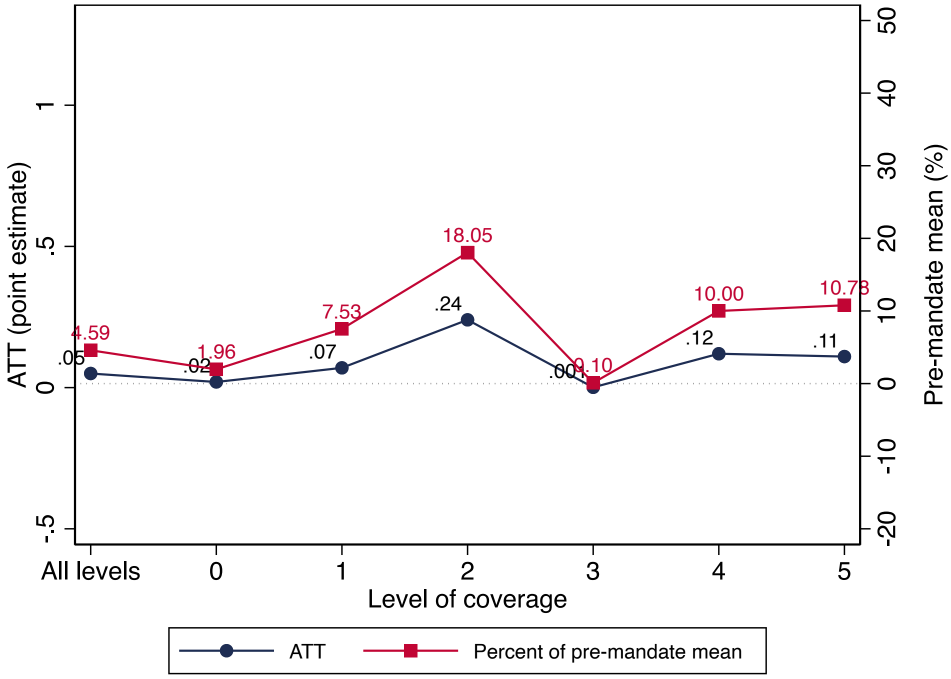
(a) All women



(b) Women 35 years and older



(c) Women under 35 years old



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

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

(a) Number of multiple births per hundred live births

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

(b) Number of infants per thousand live births

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

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

D Other dimensions of generosity and state by state GSC estimates

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

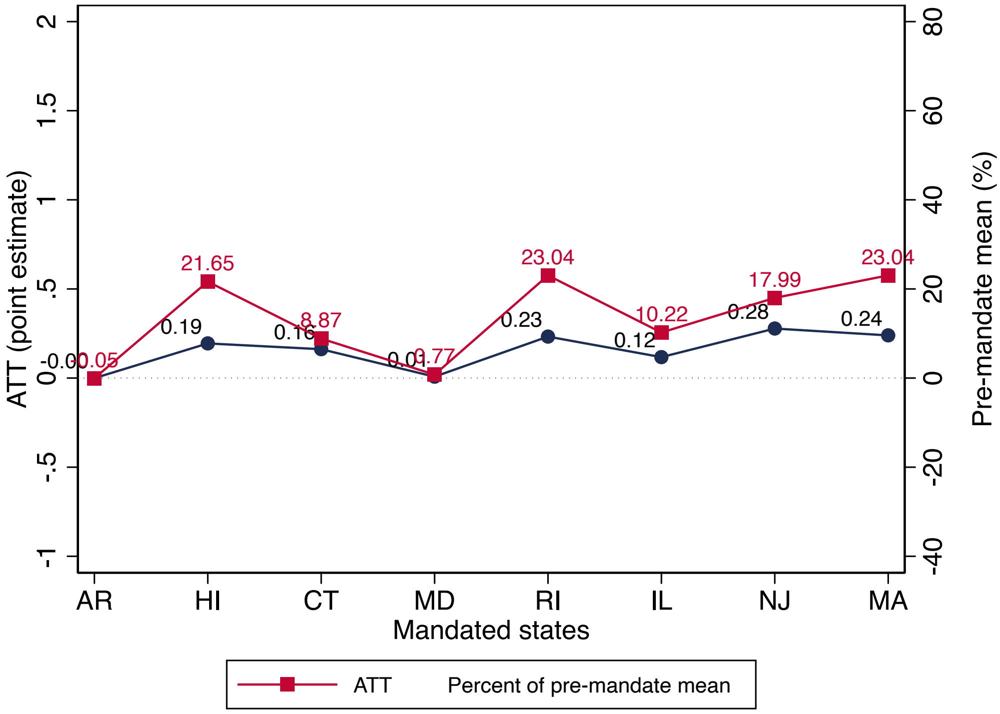
Dimension	With dimension	Without dimension	p-value of statistical significance test
	(1)	(2)	(3)
Long min infertility time	0.11	0.19***	0.00
	(0.12)	(0.07)	
Religious exemption	0.17**	0.10*	1.00
	(0.08)	(0.05)	
Firm size exemption	0.29***	0.07	1.00
	(0.09)	(0.07)	
Marital status restriction	0.08	0.21***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

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

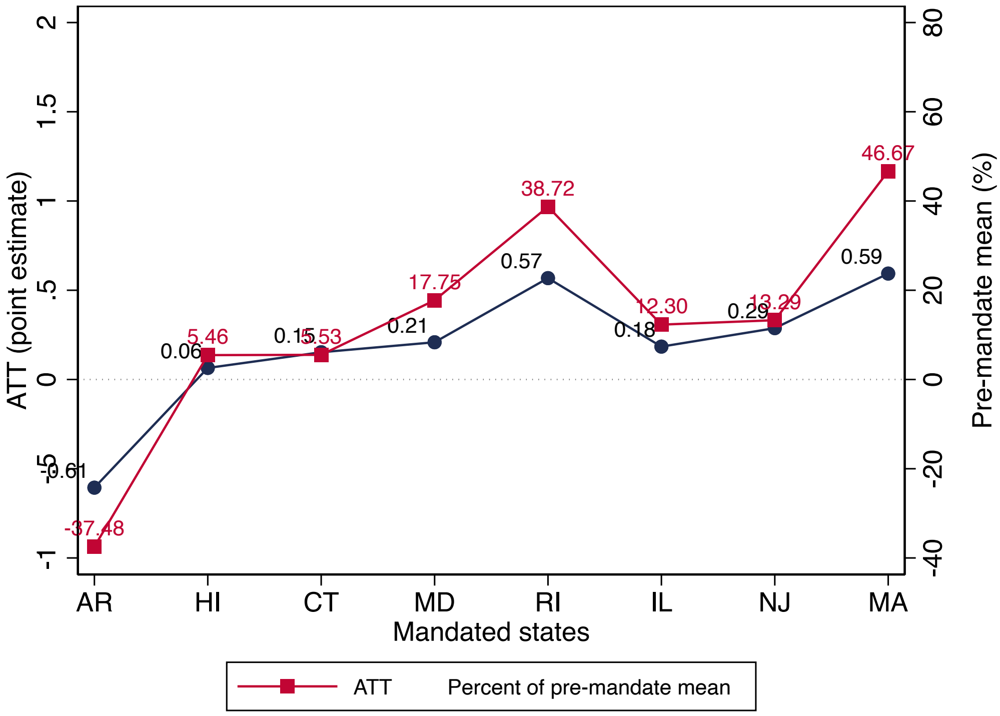
$*p < 0.10, **p < 0.05, ***p < 0.01$

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

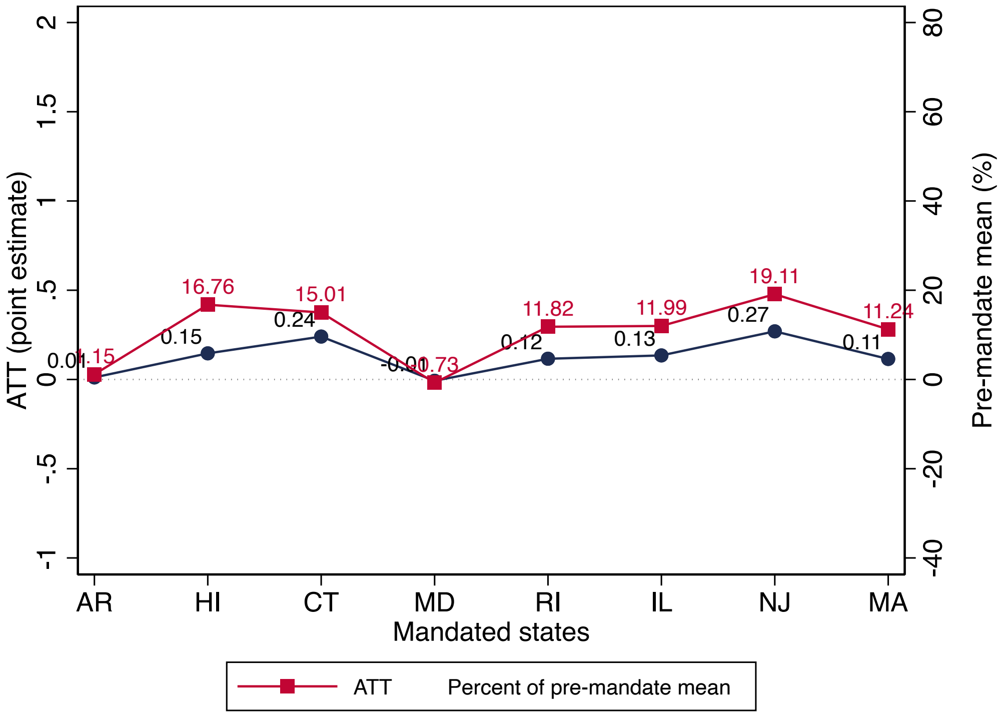
(a) All women



(b) Women 35 years and older



(c) Women under 35 years old



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

E DD and DDD estimates

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

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

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

We estimate the following equation in our DDD model:

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

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

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

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

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

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

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

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

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

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

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

(a) Multiple births per hundred live births

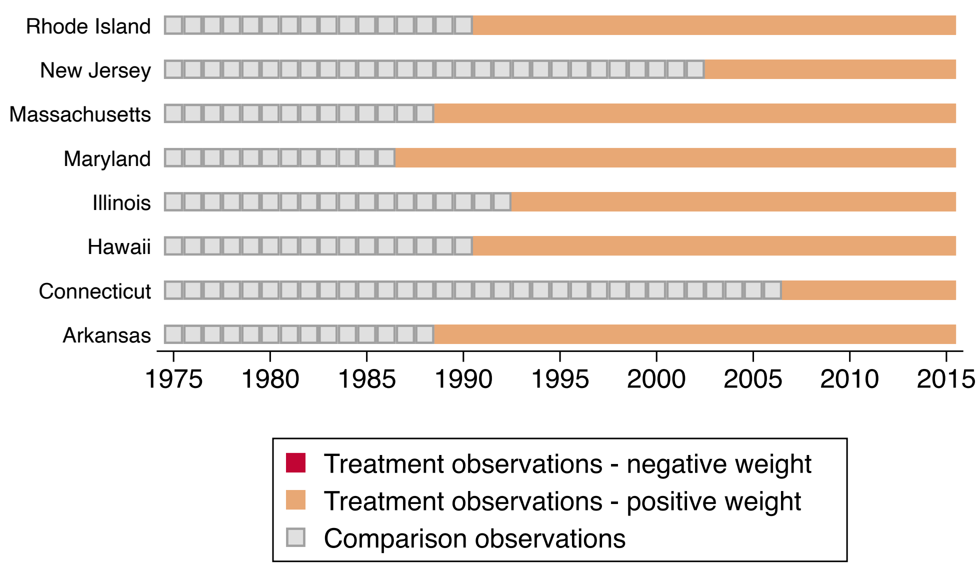
	All women		Women 35 and older		Women under 35	
	(1)	(2)	(3)	(4)	(5)	(6)
All levels	0.05	0.03	0.07	0.02	0.03	0.03*
	(0.06)	(0.02)	(0.12)	(0.07)	(0.03)	(0.01)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Covariates included	No	Yes	No	Yes	No	Yes
Observations	1,059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

(b) Number of infants per thousand live birth

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

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

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



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