Generosity of Health Insurance Coverage: Intensive and Extensive Margin Effects on Utilization*

Arezou Zaresani† Lucie Schmidt‡

April 2020

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

Abstract

We investigate patients' utilization responses to the generosity of insurance coverage for expensive medical treatments, exploring insurance mandates for in vitro fertilization treatment that vary widely in coverage generosity across US states. We find that more generous coverage increases the incidence of such treatment, and thus of potentially costly and risky multiple births. More generous coverage has intensive margin effects, reducing the number of transferred embryos per cycle. But it also has sizeable extensive margin effects, in that more older patients with lower fertility are drawn into treatment. These extensive margin effects may impose additional burdens on the healthcare system in terms of both costs and adverse health outcomes.

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

Keywords: healthcare costs, healthcare utilization, health insurance, mandated benefits, infertility treatment, in vitro fertilization

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

†University of Melbourne, IZA and TTPI, Email: azaresani@unimelb.edu.au. I acknowledge financial support from the University of Calgary, University of Victoria and University of Melbourne.

‡Williams College and NBER, Email lschmidt@williams.edu

1 Introduction

Healthcare spending in the US has risen rapidly, from 5 percent of GDP in 1960 to 17.9 percent in 2017 (CMMS, 2017). Lifestyle changes and an aging population have contributed to increases in chronic illnesses such as cancer, musculoskeletal conditions, diabetes, and heart disease. These conditions have expensive treatment options, raising concerns about access to treatment and its overall costs.

Mandated health insurance coverage for expensive medical treatments can increase accessibility by decreasing patients' out-of-pocket costs. Generous mandated coverage for a treatment can have intensive margin effects on existing patients' utilization behavior since such coverage makes seeking additional expensive treatments less costly. However, by expanding access to new patients who might have previously used a cheaper alternative, mandated coverage can also have extensive margin effects. These extensive margin effects can further contribute to increases in healthcare costs, particularly if they lead to changes in the composition of patients seeking treatment such that patients with lower probabilities of success initiate treatment. Patients' behavioral responses to the increased accessibility of expensive treatments are critical to understanding the ramifications of health policy interventions.

Mandated health insurance coverage for in vitro fertilization (IVF) treatment in the US provides an interesting case study in this context for the following reasons. First, IVF resembles other medical treatments like those for heart disease or cancer, which are expensive and have uncertain outcomes (Shapiro and Recht, 2001). Second, patients choose treatment intensity (through the number of transferred embryos) based on preferences and expected costs and benefits, and this choice directly affects both success rates and the likelihood of unintended and costly multiple births. Finally, the generosity of mandated coverage for IVF treatment 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 cycles, and mandates vary across a number of other dimensions of generosity as well.[1] This variation across states and time allows the identification of the effects of

[1] These include things like the definition of infertilty and age restrictions. We provide more in Section

generosity on utilization and outcomes.

In this paper, we empirically investigate how the generosity of mandated IVF coverage affects patients' utilization behavior and the composition of those utilizing the treatment. In the absence of data on the utilization of IVF treatment for all years, we examine multiple births as a proxy for the intensity of IVF treatment.[2] More generous coverage could have two competing effects on the likelihood of multiple births. First, holding the pool of patients constant, patients face less pressure to conceive in each cycle, and so might choose to implant fewer embryos (Jain et al., 2002; Reynolds et al., 2003). This intensive margin effect would decrease the incidence of multiple births.[3] Second, generous mandates expand access to new patients who might not have pursued treatment in the absence of insurance coverage. This extensive margin effect could lead to an increase overall in the incidence of multiple births, but might not change the number of embryos being transferred per woman or per cycle. However, these extensive margin effects could also change the composition of patients seeking treatment, such that patients with a lower probability of success initiate treatment, requiring an increase in the number of embryos transferred per woman or per cycle.[4] The overall effect of more generous coverage for IVF treatment on the incidence of multiple births is therefore ambiguous. These intensive versus extensive margin effects of IVF coverage are discussed by Bundorf et al. (2007) and Hamilton et al. (2018), but not in the context of differing generosity levels *within* the set of states which mandate coverage.

We first develop a conceptual framework to show how patients' differential fertility together with the generosity of their insurance coverage leads to differences in utilization and treatment intensity. We then use a generalized synthetic control (GSC) model to estimate the effects of IVF coverage generosity on multiple births, using Detail Natality Data on all births in the US between 1975 and 2014 and exploiting variation in generosity

2

[2]See Section 4 for a discussion of the weaknesses of this measure.

[3]Note that we are thinking of the intensive margin here in terms of intensity of the cycle. Alternatively, one could consider intensity of treatments within a given patient over time, which could show an increase in multiple births if a patient with a failed first attempt then re-enters the pool.

[4]However, if more generous coverage leads some patients to skip cheaper alternatives to IVF (for example, ovulation-boosting drugs), then the extensive margin effects could lead to a reduction in the incidence of multiple births.

across states and over time. Then, to shed additional light on intensive versus extensive margin effects, we turn to fertility clinic data from the Society for Assisted Reproductive Technologies (SART) and examine how more generous coverage affects the composition of patients, the number of initiated IVF cycles, and the number of embryos transferred per cycle. Finally, using data from the National Data Archive on Child Abuse and Neglect (NDACAN), we examine the effects of IVF generosity on adoptions of children aged 0-6, as such adoptions could be considered in some circumstances a substitute for conceiving through IVF.

Our empirical analysis has three main findings. First, we find that more generous coverage of IVF treatment increases the incidence of multiple births. After controlling for state-level characteristics, we find that multiple births are 26.9% higher in states with the most generous coverage relative to states with no mandated coverage. Second, we find evidence of intensive margin effects of generosity for all women, where states with the most generous coverage have fewer embryos transferred per cycle. Finally, we find that the states with the most generous coverage also see significant increases in the share of cycles performed on older women with lower fertility, as new patients are drawn into treatment. These compositional extensive effects dominate the decrease in embryos transferred per cycle, leading to overall increases in multiple births.

Our findings suggest that extensive margin effects are important to understanding the policy implications of increased health insurance generosity, consistent with previous work on the role played by incentives in healthcare utilization. Chernew et al. (2000) suggest that in an optimal insurance plan patients should pay higher out-of-pocket costs for more expensive treatment. Einav et al. (2016) (in the case of breast cancer treatments) and Hamilton et al. (2018) (in the case of infertility treatments) both suggest that top-up pricing for more aggressive treatments could be optimal.

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. Treatment for infertility usually begins with medical tests and physician advice, and is often followed by the woman's use of one of several drugs to stimulate egg production. If these less expensive treatment methods are not successful, then assisted reproductive technologies such as in vitro fertilization (IVF) are often recommended. Success rates of a single IVF cycle are as low as 20 percent (CDC, 2015), and many patients require more than one cycle of treatment to achieve a live birth. The costs of one cycle of IVF can be as high as 46 percent of the average US family's annual disposable income (Kissin et al., 2016).

In IVF, eggs are extracted, a sperm sample is obtained, and eggs and sperm are then manually combined. The fertilized eggs, called embryos, are then transferred into the woman's uterus.[5] The Practice Committee of the American Society of Reproductive Medicine provides guidelines on the maximum number of embryos to transfer per cycle (Klitzman, 2016).[6] However, given the high costs and low success rates of IVF, patients often wish to exceed these guidelines to improve their odds of success, and in doing so increase the likelihood of multiple births. Although multiple births are more costly and risky for both mothers and infants, most monetary costs are covered by insurance, and many patients with fertility problems view multiple births as a desirable outcome (Gleicher and Barad, 2009),

2.2 Mandated IVF coverage in health insurance plans

Due in part to concerns about the high cost of IVF treatment, between 1978 and 2014 15 US states passed legislation pertaining to coverage of infertility treatment in employer

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

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

provided private health insurance plans.[7] In these *mandate to cover* states, private health insurance companies are required to cover infertility treatment in all of their policies.[8]

The level of coverage in mandate to cover states is quite heterogeneous. During our sample period, Montana, New York, Ohio, and West Virginia mandate coverage for some types of infertility treatment, but do not require coverage of IVF. Arkansas and Hawaii mandate coverage for only one cycle of IVF; Connecticut mandates up to two; Rhode Island and Maryland mandate up to three; Illinois and New Jersey mandate up to four; and Massachusetts has no limit.[9] Mandates also vary along a number of other dimensions, including (but not limited to) age restrictions, coverage of unmarried women, minimum years of infertility to qualify for coverage, and whether mandates apply to health maintenance organizations (HMOs). However, as shown in Table 1, these dimensions of generosity are highly correlated with the mandated number of cycles, so we treat the number of cycles as a proxy for the overall generosity level of mandated coverage. There are 35 states which never legislated policies to mandate coverage for infertility treatments. These *never mandate states* serve as a control group in our analysis.

2.3 Previous work

Our paper is related to the literature investigating the effects of mandated coverage for infertility treatment on a variety of outcomes including utilization of treatment, infant and child health outcomes, fertility, age at first birth, time of marriage, women's choice to pursue professional careers, and labor supply over the life cycle (Schmidt, 2005; Bitler and Schmidt, 2006; Bundorf et al., 2007; Schmidt, 2007; Bitler, 2008; Bitler and Schmidt, 2012; Buckles, 2013; Abramowitz, 2014; Machado and Sanz-de Galdeano, 2015; Kroeger and La Mattina, 2017; Abramowitz, 2017; Gershoni and Low, 2020a,b). Most of these studies use either state-year or state-year-age variation in mandated IVF coverage in (re-

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

[8]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.

[9]Since the end of our sample period in 2014, three additional states have mandated IVF coverage: Utah, Delaware, and New York. We do not include these mandates in our analysis.

spectively) difference-in-difference (DD) and difference-in-difference-in-differences (DDD) frameworks.[10]

The studies that relate most closely to our work are those studies that examine the effects of the state mandates on multiple births.[11] Studies that examine the effects of mandated coverage tend to find that it increases multiple births. Buckles (2013) finds that "strong" mandate to cover laws (those that include coverage for IVF and apply to most private firms) have a small but statistically insignificant impact on the incidence of multiple births. Bitler (2008) shows that the mandates are associated with higher rates of multiple births and worse health outcomes in terms of birth weight and gestation. Bundorf et al. (2007) show that the mandates are associated with a significant increase in multiple births per delivery. However, studies that use clinic-level data find that treated patients with health insurance plans covering IVF treatment transfer fewer embryos compared to those with no insurance coverage (Jain et al., 2002; Reynolds et al., 2003; Henne and Bundorf, 2008; Hamilton and McManus, 2012). Much of this previous work ignores the differences in generosity within the set of states that mandate coverage for IVF.

2.4 Our contribution

Our contribution to this literature is twofold. First, we study how patient utilization responds to the *generosity* of mandated IVF coverage. This is important for understanding the cost implications, since more generous coverage could affect utilization on both the intensive and extensive margins, and therefore could alter the composition of those seeking treatment. The second contribution is methodological. The DD approach used in much of the previous literature relies on the assumption that trends in the treatment (mandate to cover) and control (never mandate) states would have evolved in the same way in the absence of the mandates. While previous papers all address this parallel trends assumption when comparing all mandate states to nonmandate states, it might be less

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

[11] Bhalotra et al. (2020) find that a Swedish single embryo transfer policy reduced multiple births and improved maternal and infant health.

plausible in the context of differences in generosity within the set of states that choose to mandate coverage of IVF. Given this concern, we use a GSC model, described in Section 5, to generate causal estimates. However, we also estimate DD (exploiting variation across states and time) and DDD (exploiting variation across states, time and women's age) models as robustness tests, and for easier comparability to the previous literature.

3 Conceptual framework

In this section, we build upon Hamilton and McManus (2012) to provide a conceptual framework to illustrate how the generosity of IVF coverage in a patient's insurance plan and her underlying fertility affect her utilization behavior.

Many factors affect a patient's decision to initiate IVF treatment, as well as her choice of treatment intensity. First, as noted previously, per-cycle costs of treatment are quite high. In addition, most patients need more than one cycle of treatment to achieve in a live birth. Therefore, patients face strong financial incentives to minimize their total treatment costs by conceiving in fewer cycles. As a result, they may transfer more embryos in a given cycle in order to improve their chances of success, despite the fact that this also increases the likelihood of multiple births. Health insurance plans with more generous IVF coverage reduce the out-of-pocket costs of the treatment and potentially reduce the pressure to conceive in fewer cycles. As a result, patients with more covered cycles might decide to transfer fewer embryos in a given cycle, subsequently reducing the probability of multiple births.

Patients get utility from consumption (α) and having children (b). Each patient is endowed with fixed income I. A patient makes a decision d from the options of natural conception (N), initiating infertility treatment (IVF) and adopting an infant (A) to maximize their utility defined as:

$$\max_{d \in \{N, IVF, A\}} U(\alpha, b) = \alpha + v_d(b), \tag{1}$$

where $v_d(.)$ denotes the utility associated with choice d, and b denotes the number of infants resulting from choice d. We assume that, conditional on a successful birth, patients prefer fewer infants resulting from a delivery. We therefore assume that $v_d(.) > 0$ and $v'_d(.) < 0$ for all choice of d. All patients prefer to have their own biological infant with a natural conception where $v_N(.) > v_{IVF}(.)$ and $v_N(.) > v_A(.)$.

Patients' consumption (α) is their income net of the cost of their choice d, defined as $\alpha = I - p_d$ where p_d is the cost associated with decision d. The cost of natural conception p_N is assumed to be zero. The cost of adoption is also assumed to be fixed at $p_A > 0$. The costs of IVF treatment consist of two parts; the fixed costs of initiating a treatment ($\eta > 0$) and the per cycle costs which might be covered by an insurance plan ($\lambda > 0$) which is defined as $p_{IVF} = \eta + \lambda$.

The more embryos transferred, the higher the probability of conceiving an infant, and the greater the likelihood of a multiple birth. We assume that the number of infants resulting from a cycle of treatment is $b = k\kappa$, where $k \in [1, \bar{k}]$ is the number of transferred embryos and κ is a fixed parameter denoting the probability of conceiving multiples naturally. $\bar{k}$ denotes the maximum number of implanted embryos recommended by professional guidelines. We also assume that the number of transferred embryos (k) is a function of a patient's fertility $f \in [\bar{F}, \underline{F}]$ and the number of the cycles covered in their insurance plan ($\bar{r}$) and defined as $k = g(f, \bar{r})$. $\bar{F}$ and $\underline{F}$ denote the fertility of patients with respectively low and high chances of natural conception. Patients with lower fertility face incentives to transfer more embryos ($g_f(.) < 0$), and patients with more covered cycles transfer fewer embryos ($g_r(.) < 0$). Patients' fertility f and their insurance coverage for IVF treatment $\bar{r}$ are the only sources of heterogeneity in our model. For simplicity and with no loss of generality, we assume $k = \frac{\bar{F}\bar{k}}{f\bar{r}}$.

We assume that the per-cycle probability of conceiving an infant depends on the number of transferred embryos defined as $\phi(k)$, where more transferred embryos improve the chances of success ($\phi_k(.) > 0$). The probability of a natural conception is denoted by γ, and we assume $\phi(k) = \gamma k$.

Figure 1 illustrates patients' choice by their fertility status f. Patients with higher $f \in (\frac{1}{\gamma}, \overline{F}]$ are more likely to naturally conceive an infant. Patients with lower $f \in [\frac{1}{\gamma\bar{r}\bar{k}}, \frac{1}{\gamma}]$

would use IVF treatment. When the number of covered cycles ($\bar{r}$) increases, more patients with lower fertility would choose initiating IVF treatment over adopting a child. These patients would increase their chances of conceiving an infant by implanting more embryos which might result in a multiple birth. Patients with $f \in [\underline{F}, \frac{1}{\gamma \bar{r} k})$ are predicted to opt for adoption rather than to initiate IVF treatment.

This framework is static, and the true optimization problem faced by a patient is clearly dynamic in nature. However, if we assume that $\bar{r}$ denotes the number of *remaining* covered cycles, then we can think of this model as representing a given stage of a patient's dynamic decision making process.

4 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 mother's age, education, and race, father's race, parental marital status, and state of residence; and infant information such as sex, birth order, and plurality (single or multiple birth). Our study sample includes the 12 mandate to cover states (treatment group) and the 35 never mandate states (control group). We aggregate the data into state-year cells for our empirical analysis.[12]

Our primary outcome variable is the multiple birth rate defined as the number of multiple births (i.e. not singletons) per hundred live births.[13] Multiple births are a useful proxy for the aggressiveness of treatment. More than one-third of twins and more than three-quarters of triplets and higher order multiples in the US in 2011 resulted from conception assisted by infertility treatments (Kulkarni et al., 2013). They can also be

[12] We use the NBER data files. Public use data include mother's state of residence only through 2004, so we use restricted access data files from 2005-2014. We impute missing values in the state-year aggregated data by setting them to the average of the corresponding variable in the years before and after.

[13] 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 1 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.

costly and risky for both mothers and infants.[14] However, one caveat of this approach is that in the birth certificate data, we have no way of knowing whether the multiple births are naturally occurring, or due to IVF, or due to other infertility treatments besides IVF. Our multiple birth indicator also cannot differentiate between 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 births.

Second, we use the March Annual Social and Economic Supplement of the Current Population Survey (CPS) to create control variables at the state-year level, including the population percentage of women of childbearing age, the female labor force participation rate and real per capita income.[15] 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 +500 employees) as a proxy for the share of workers in self-insured firms and therefore not subject to the mandates under ERISA.[16]

Third, we use clinic-level data collected from 1996 to 2010 by SART to study patients' utilization of IVF treatment.[17] The data include information on the number of cycles initiated in each clinic, the share of cycles performed on women 35 and older, and the average number of embryos transferred by mothers' age.

Finally, we use data on child adoptions from the National Data Archive on Child Abuse and Neglect (NDACAN) from 2000 to 2014.[18] We create a variable to represent the number of young adopted children (ages 0-6) per one thousand live births in that state and year. Our data do not include private adoptions (either domestic or international). However, our analyses of the effects of the insurance mandates will be biased only if the

[14]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 rate of caesarean sections. 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).

[15]We use the NBER files, and convert all dollar values to 2007 dollars using the Consumer Price Index.

[16]It has been shown that large firms are more likely to self-insure (Gabel et al., 2003; Park, 2000).

[17]SART is a voluntary reporting system and does not regulate clinic practices. About 10% of clinics do not report data. We exclude frozen and donor cycles, since only fresh and non-donor cycles are covered by mandates in many states.

[18]This is a federally mandated data collection system for all children in foster care and on children adopted under the auspices of the state public child welfare agency.

generosity of mandated IVF coverage differentially affects private adoptions versus those through the state welfare system.

5 Identification strategy

States mandated insurance coverage of IVF at different times. We could follow the previous literature and use this state- and time-level variation to estimate the effects of the generosity of mandated coverage on the incidence of multiple births using a DD framework. However, interpretation of estimated effects as causal requires that in the absence of treatment, the incidence of multiple births in the treated and control states would have followed parallel paths over time. Figure 3 plots trends in the incidence of multiple births by the generosity level of the mandated coverage. These figures suggest that the parallel trend assumption might be violated in our context.[19]

In order to estimate causal effects when the parallel trends assumption is likely to be violated, we use a GSC framework (Xu, 2017). We estimate a model of the form:

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

where i and t respectively denote state and time. y_{it} denotes the outcome variable in state i at year t. Our main outcome variables are the multiple birth rate (the number of multiple births per hundred live births) and the number of infants per thousand live births. D_{it} is a dummy variable which is coded as 1 for treated state i in years following the mandated coverage. We estimate the model separately for each generosity level indicated in Table 1. We follow Schmidt (2007) and allow mandated coverage to affect multiple births with a two year delay. This effective mandated coverage year accounts for two factors: first, infertility treatments may not lead immediately to a conception, and second, a successful conception will not translate into a birth until nine months later.

The vector X_{it} is a set of time-varying state-level characteristics which includes mothers' age, marital status, and education, mothers' and fathers' race, infant's sex, birth

[19]Similar figures for the number of infants per one thousand live births are available from the authors by request.

weight, and five-minute APGAR score. We also include the state-level socioeconomic characteristics from the CPS discussed above.

λ_i is a $r \times 1$ vector of state-specific intercepts. f_t is a $r \times 1$ vector of time-varying coefficients which captures unobserved common factors. r is the estimated number of confounding factors. The factor component of the model, $\lambda_i' f_t$ covers a wide range of unobserved heterogeneity. It absorbs all unobserved confounders that can be decomposed into a state-year multiplicative form (i.e. $U_{it} = a_i \times b_t$), but it does not capture unobserved confounders that are independent across states. ϵ_{it} captures any remaining unobserved components that affect the incidence of multiple births.

The coefficients of interest are δ_{it}. The average treatment effect on the treated (ATT) at time t is $\widehat{ATT}_t = \frac{1}{|Treated|} \sum_{i \in Treated} \delta_{it}$, where $Treated$ denotes treated states. We use data from a 15-year window around the effective mandated coverage year (15 pre- and 15 post-treatment periods).[20] We estimate standard errors with a parametric bootstrapping procedure using 2,000 re-sampling draws of the residuals (Xu, 2017).

A generalization of the conventional fixed effects model, the GSC approach uses a linear interactive fixed effects model to impute the treated counterfactuals using the information from the treatment group in pre-treatment periods and the control group, in the spirit of the weighting scheme of the original synthetic control method (Abadie et al., 2010).[21] The number of confounding factors (r above) is chosen within a data-driven cross-validation procedure. A more flexible interactive fixed effect (i.e. a larger r) covers a wider range of unobserved heterogeneity. We provide more details on the estimation strategy of a GSC framework in Appendix A.

The main advantage of using a GSC framework instead of a DD framework is that a GSC framework provides causal estimates when the parallel trends assumption is likely

[20] 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.

[21] The GSC framework has several advantages relative to the original synthetic control developed by Abadie et al. (2010). First, it allows for more than one treated state with variable treatment periods. Second, the GSC framework provides estimates of standard errors and confidence intervals, making inference more reliable. Third, it provides a data-driven procedure to select the right number of factors in an interacted fixed effect model and reduces the risk of over-fitting. This approach furthermore enables us to take advantage of the long pre-treatment panel to decrease the bias of the estimated effects. See (Abadie, 2019) for a review of recent synthetic control methods.

to be violated.[22] A GSC framework also allows for more flexible state-year interactions (i.e., state specific time trends) to absorb a wider range of unobserved heterogeneity than a DD model allows. A GSC framework allows the data to tell us which model is more appropriate by estimation of r. For specific estimated values of r, a GSC framework reduces to a DD framework.[23]

6 Results

6.1 Descriptive evidence

Table 2 presents summary statistics from birth certificate data from 1975-2014, presented in ten year intervals and broken out by IVF mandate status. Mothers in more recent years are on average older, more educated, and less likely to be married. Multiple births per hundred live births and the number of infants per thousand live births are also higher in more recent years, reflecting an increase in infertility treatments as well as an increase in the share of older mothers (older women are more likely to have multiple births even in the absence of infertility treatment). 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 2 plots the multiple birth rate by women's age. The incidence of multiple births increases by women's age, and this pattern is stronger in recent decades. The age of 35 is considered to be 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 versus women younger than 35 years.

Figure 3 plots trends in the number of multiple births per hundred live births by generosity level of mandated IVF coverage, first for all women, then separately for older

[22] A DDD framework can also be used for estimating causal effects when the parallel trends assumption is likely to be violated, since it adds a third dimension (in our case, mother's age in addition to state and year). We estimate DDD models as a robustness check.

[23] 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 the conventional DD model with state and time fixed effects.

and younger women. Three main patterns emerge. First, the incidence of multiple births is increasing across all states over our sample period. Second, more generous coverage is associated with more rapid growth in the incidence of multiple births. Third, the association between coverage generosity and the incidence of multiple births is stronger for older women than for younger women.

6.2 Estimation results from the GSC model

Plots presenting the estimated counterfactual and estimated effects on the treated (mandate to cover) states for each level of coverage for all women are presented in Figure 4.[24] The plots 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.[25] Table 5 presents the estimated effects of the generosity level of mandated coverage on the number of multiple births per hundred live births, using the GSC framework specified in Equation (2). 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. The first column shows that any mandated coverage increases the multiple birth rate by 0.10 percentage points relative to the never mandated states, approximately 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 increase in the multiple birth rate (or a 4.42% increase).

Panels B through G show the estimated effects broken out by the level of generosity. Panel B shows that coverage for less invasive infertility treatment only (level 0) has no effect on 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, in general, states with more generous IVF coverage exhibit larger increases in multiple birth rates. Estimated effects with covariates range from a 0.08 percentage point increase (8.33%) in states with level 1 coverage to a 0.28 percentage point increase (26.92%) in states with

[24]Figure 5 and Figure 6 present the corresponding estimates respectively for women 35 and older versus women below 35 years.

[25]Similar figures for the number of infants per one thousand births are available from the authors by request.

level 5 coverage.

The remaining columns of Table 5 present the estimates for women 35 and older versus younger than 35. After controlling for covariates, the estimated effects for women 35 and older tend to be larger than the effects for younger women, especially at higher levels of coverage. For older women, the estimated effects after controlling for state-level characteristics vary from -0.24 percentage points (-17.26%) in states with level 1 coverage to 0.56 percentage points (44.09%) in states with level 5 coverage. The estimated effects for younger women are much smaller and range from 0.07 percentage points (7.52%) in level 1 states to 0.21 percentage points (20.59%) in level 5 states. These findings are consistent with our conceptual framework presented in Section 3.

While the multiple birth rate simply tells us whether the birth included more than one infant, our alternative outcome measure, the number of infants per thousand births, allows, for example, triplets to count more than twins. Table 6 presents the effects of the generosity level of mandated coverage on the number of infants per thousand live births. The overall findings are quite consistent with those from the multiple birth rate specification. The estimated effect of any mandated coverage (Panel A) after controlling for covariates is 0.64 infant (or, a 5.51% increase in the number of infants per thousand live births). The estimated effects by the generosity level of coverage after including covariates range from 0.91 infant (9.37%) in states with level 1 coverage to 2.92 infants (27.68%) in states with level 5 coverage, and again, the effects are larger for older women.[26]

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.

As a robustness check, and to facilitate comparison with the previous literature, we

[26]There are other dimensions besides age that are strongly associated with infertility and IVF coverage, including education, marital status, and race (Bitler and Schmidt, 2006). College educated women face incentives to postpone childbearing and invest in their professional careers, and are also more likely to work in jobs that offer private health insurance. Married women struggling with fertility seek infertility treatment and especially IVF more often than unmarried women, and some mandated coverage explicitly excludes unmarried women. Although white women are less likely to experience infertility than black women, they are more likely to seek infertility treatment. We have estimated the results from Tables 5 and Table 6 separately along these dimensions, and results are largely consistent with the patterns found in the previous literature. Results are available from authors on request.

also estimate the effects of the mandated coverage on the incidence of multiple birth using DD (exploiting variation across states and over time) and DDD (exploiting variation across states, over time, and by age cohort) frameworks. Appendix Table B.1 presents the estimated effects from these specifications for multiple births per hundred live births, and and Appendix Table B.2 presents similar results for the number of infants per thousand live births. The overall story is the same as our findings from the GSC framework; more generous coverage is associated with an increase in the incidence of multiple births, and this association is stronger for older women.

7 Intensive and extensive margin effects

Our GSC results show that more generous coverage leads to an increase in the incidence of multiple births. This is in spite of speculation that more generous coverage might reduce multiple births by reducing the incentives to transfer more embryos per cycle. If mandated coverage has only intensive margin effects on patients' utilization behavior, then we would expect more generous coverage to decrease the incidence of multiple births. However, more generous IVF coverage could also have extensive margin effects, as new patients with lower probabilities of success seek treatment.

To disentangle the intensive and extensive margin effects of the generosity of mandated coverage on the incidence of multiple births, we use two additional datasets. First, we investigate patients' utilization behavior using fertility clinic-level data. Second, we investigate child adoption as the main alternative to a live birth. However, since both of these data sets begin after several of the mandates are passed, these analyses should be thought of as descriptive and not as providing causal estimates.

7.1 Evidence from IVF clinics

We turn to SART clinic-level data from 1996 to 2010 to directly investigate whether the generosity of mandated coverage affects IVF utilization. Table 3 presents summary statistics, and Figure 7 plots trends in our outcomes by coverage generosity. The average number of embryos is decreasing over our sample period in both treatment and control states, likely due in part to changes in medical recommendations.[27] More embryos are transferred per cycle for women 35 and older than for younger women. The share of cycles performed on women 35 and older is 10 percentage points higher in recent years for treatment states relative to control states.

We use a linear mixed effects (ME) model to investigate the relationship between coverage generosity and patients' utilization behavior. Our ME model exploits random variation between clinics within states in addition to the variation across states. We are unable to use GSC or DD models here, because the mandated coverage date for 5 out of 8 states falls before the availability of SART data. Including clinic or state fixed effects would absorb all the variation.[28] An ME model utilizes the hierarchical structure of the data where the observations (i.e. clinics) are nested in groups with particular characteristics (i.e. states with various levels of coverage). We estimate a model specified as:

$$y_{ist} = \alpha_0 + \alpha_1 Level_{st} + \alpha_2 X'_{st} + \lambda_t + \nu\gamma_i + \omega\gamma_s + \epsilon_{ist}, \tag{3}$$

where i, s and t respectively denote clinic, state, and year. y_{ist} denotes the outcome variable in clinic i in state s at year t. Our outcome variables include the total number of cycles, the share of cycles performed on women 35 and older, and the average number of transferred embryos per cycle. $Level_{st}$ is an indicator for the generosity level of mandated coverage in state s at year t, with never mandated states as the control group. The vector X_{st} includes the same time-varying state characteristics used in our GSC analysis. λ_t denotes year fixed effects, which pick up any factors changing over time that are common across states and clinics (e.g., advances in technology at the national level). γ_i and γ_s respectively denote clinic and state random effects. ϵ_{ist} captures any remaining unobserved factors affecting the outcome variable. The parameter of interest is α_1, which

[27] A major change to SART guidelines occurred in 2004. We estimated an event study model, and found that this change was associated with a reduction in the number of embryos transferred for both younger and older patients. However, the estimated effects did not vary by coverage generosity. Results are available from authors on request.

[28] 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).

captures the relationship between the the generosity level of mandated coverage and our outcome variables.

Table 7 presents the estimated effects for all women as well as results broken out by age. These results suggest the following: First, more generous coverage is associated with a significant increase in the share of cycles initiated by older women, which is suggestive of extensive margin effects on the composition of the pool of patients. Given that older women transfer more embryos per cycle, this would imply an increase in multiple births. Second, the relationship between the generosity of mandate coverage and the average number of transferred embryos per cycle reflects intensive margin effects: more generous coverage is associated with fewer transferred embryos for both older and younger women, with stronger effects for younger women. This would imply a decrease in multiple births. The fact that our GSC results using birth data show an overall increase in multiple births suggests that the compositional or extensive margin effect must dominate.

7.2 Evidence from child adoption

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

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

We use NDACAN's child adoption data for children aged 0-6 years from 1995 to 2014 to investigate the relationship between the generosity level of mandated IVF coverage and child adoption.[29] Table 4 presents descriptive statistics for these data. In the early years of our sample period, the adoption rate was higher in our treatment states than in the control states. However, by the later half of our time period, this pattern had reversed itself, so that the never-mandated states saw two more adopted children per one thousand live births than did the treatment states. Figure 8 plots the number of adopted children per thousand newborn infants by the age of the adoptive mother. This figure provides two main insights. First, more generous coverage is associated with lower rates of child adoption. Second, adoption rates are higher for older women than younger women.

To examine the effects of generosity of IVF coverage on child adoption, we estimate an ME model similar to Equation (3), including time fixed effects and state random effects.[30] Table 8 presents the estimated effects, first for all women and then broken out by the age of the mother. Our results suggest a negative association between the generosity level of mandated coverage and the number of adopted children per thousand newborn infants that is much stronger for older women than for their younger counterparts.

Our analyses of these three different data sources (birth certificate data, data from IVF clinics, and adoption data) have three main takeaways. First, more generous IVF coverage increases the incidence of multiple births. Second, the intensive margin effects of more generous coverage (i.e. the decrease in the number of transferred embryos) are found for all women, but are stronger for younger patients than older patients. Third, the extensive margin effects of more generous coverage show a compositional change, where the share of cycles performed on women over 35 increases with coverage generosity. This is mirrored

[29] We focus on these ages since younger children might be closer substitutes for newborn infants.

[30] Similar to our analysis of SART data, we are unable to use GSC or DD models because the mandated coverage date for 5 out of 8 mandated states falls before the availability of the adoption data.

by a decrease in child adoption to older women in states with more generous coverage. These findings suggest that the extensive margin effects of more generous coverage are stronger than the intensive margin effects on the number of embryos transferred per cycle, resulting in the overall increase in risky and costly multiple births.

8 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 and find that more generous coverage significantly increases the incidence of risky and expensive multiple births. This is true despite the fact that more generous coverage has been proposed as a way to decrease the incidence of multiple births by affecting patients' utilization behavior along the intensive margin, i.e. by encouraging less intensive treatment through the transfer of fewer embryos per cycle. We find that while more generous coverage has these predicted intensive margin effects for all women (and stronger effects for younger women), it also has sizeable extensive margin effects, increasing the share of cycles performed on older women. Our analysis highlights the importance of extensive margin effects of increased accessibility of expensive medical treatment through insurance coverage.

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 detection of pre-cancers. However, they also find that a large share of the increased screenings resulted from mandates that were not consistent with recent American Cancer Society 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/limiting the aggressiveness of treatments; or for imposing a "top-up" price for more expensive treatments; or some combination of the two. In the IVF context, Hamilton et al. (2018)

argue that a “value-based” policy in which insurance plans cover single embryo cycles but patients must pay a top-up cost for transferring additional embryos could maximize welfare. Ignoring compositional extensive margin 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. *Review of Economic Studies*, 72(1):1–19.

Abadie, A. (2019). 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.

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

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

Bitler, M. P. (2008). Effects of increased access to infertility treatment on infant and child health outcomes: Evidence from health insurance mandates.

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.

CDC (2015). Centers for Disease Control (CDC) 2015 Fertility Clinic Success Rates Report. https://www.cdc.gov/art/artdata/index.html.

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

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

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

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

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

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

Gabel, J. R., Jensen, G. A., and Hawkins, S. (2003). Self-insurance in times of growing and retreating managed care.

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

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

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.*

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

Hamilton, B. H., Jungheim, E., Mcmanus, B., and Pantano, J. (2018). Health care access, costs , and treatment dynamics : Evidence from In Vitro fertilization. *American Economic Review*, 108(12):3725–3777.

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

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

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

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

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

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

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

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

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.

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).

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

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

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

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

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

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

Tables

Table 1: Mandated infertility coverage in private health insurance plans

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

Source: RESOLVE: The National Infertility Association http://www.resolve.org/family-building-options/insurance_coverage/state-coverage.html [Accessed on June 15, 2017] and National Conference of State Legislatures (NCSL) http://www.ncsl.org/research/health/insurance-coverage-for-infertility-laws.aspx [Accessed on November 29, 2019] and https://www.reproductivefacts.org/resources/state-infertility-insurance-laws/ [Accessed on November 2019].

Table 2: Summary statistics, 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 4 are used to calculate statistics in this table. Standard deviations appear in parentheses.

Table 3: Summary statistics for SART reporting clinics, 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 4: 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 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 5: Effects of IVF coverage generosity level on multiple births per hundred live births, GSC model

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

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

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

Table 6: Effects of IVF coverage generosity level on the number of infants per thousand live births, GSC model

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

Note: See notes for Table 5

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

Table 7: Effects of IVF coverage generosity level on patients' IVF utilization behavior, ME model

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

Notes: Source: SART data of all women receiving IVF in a clinic in the US from 1996 to 2010. All estimates include year fixed effects and clinic random effects. Included state-level covariates from the CPS are listed in Notes to Table 5. We also control for the number of IVF clinics in each state. Standard errors are clustered at the state level and appear in parentheses.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 8: Effects of IVF coverage generosity level on adopted children per thousand live births, ME model

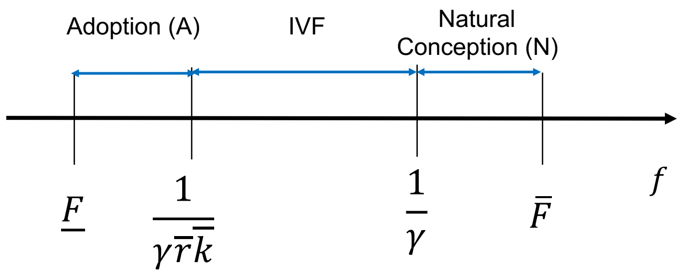
	All women				Women 35 and older				Women under 35			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
All levels	0.62	0.20			-3.12	-1.64			0.00	-0.00		
	(0.66)	(0.67)			(5.23)	(5.17)			(0.15)	(0.13)		
Level 0			1.42	0.73			17.31	16.00			-0.06	0.06
			(1.63)	(1.54)			(17.95)	(14.17)			(0.52)	(0.29)
Level 1			0.35	-2.98			2.78	-14.19			0.08	-0.66
			(1.10)	(2.02)			(5.66)	(21.33)			(0.14)	(0.66)
Level 2			1.86***	1.47**			1.75	6.60*			0.14	0.15
			(0.44)	(0.64)			(3.22)	(3.94)			(0.10)	(0.16)
Level 3			0.71	0.41			-6.88	-15.15			-0.41	-0.10
			(2.87)	(2.50)			(15.43)	(15.77)			(0.54)	(0.31)
Level 4			-0.41	-0.31			-12.94***	-6.45			0.05	0.09
			(0.43)	(0.75)			(3.79)	(5.44)			(0.16)	(0.12)
Level 5			-0.68	-0.28			-21.36***	-22.68***			-0.49***	-0.22
			(0.49)	(0.77)			(3.21)	(6.40)			(0.14)	(0.17)
Constant	2.05***	14.14	2.02***	14.92	19.08***	-47.27	17.69***	-54.02	0.64***	-2.14	0.67***	-2.18
	(0.49)	(14.07)	(0.52)	(14.52)	(3.99)	(138.19)	(4.04)	(141.14)	(0.17)	(5.00)	(0.18)	(4.96)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State random effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Covars	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Number of state-year cells	927	883	927	883	927	882	927	882	897	875	897	875

Note: Data include children ages 0-6 adopted between 1994 to 2014, aggregated into state-year cells. All estimated effects include year fixed effects and state random effects. Included state-level covariates from the CPS are listed in Notes to Table 5. We also control for the number of IVF clinics in each state. Robust standard errors appear in parentheses.

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

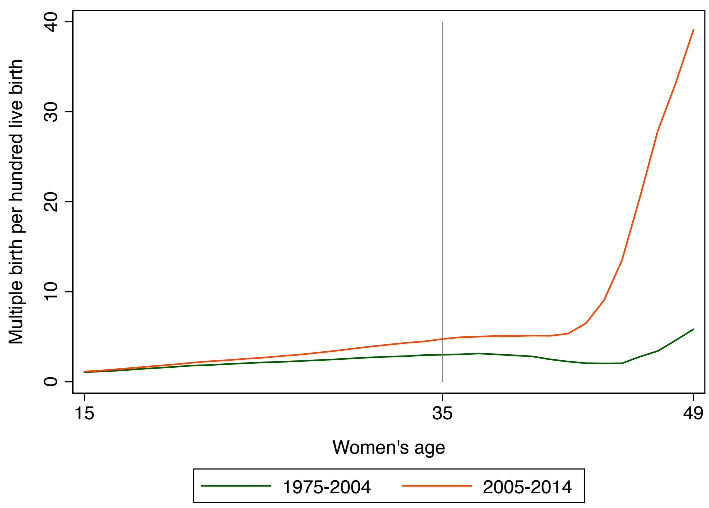
Figures

Figure 1: Patients' treatment decision by their fertility level



Note: This figure presents patients' choices for adopting an infant, IVF treatment and natural conception by their fertility (f) and the number of IVF cycles covered in their heath insurance plan ($\bar{r}$). $\bar{F}$ and $\underline{F}$ respectively denote the upper and lower limits of natural fertility. γ denotes the probability conceiving an infant naturally. $\bar{k}$ denotes the maximum number of embryos that can be implanted.

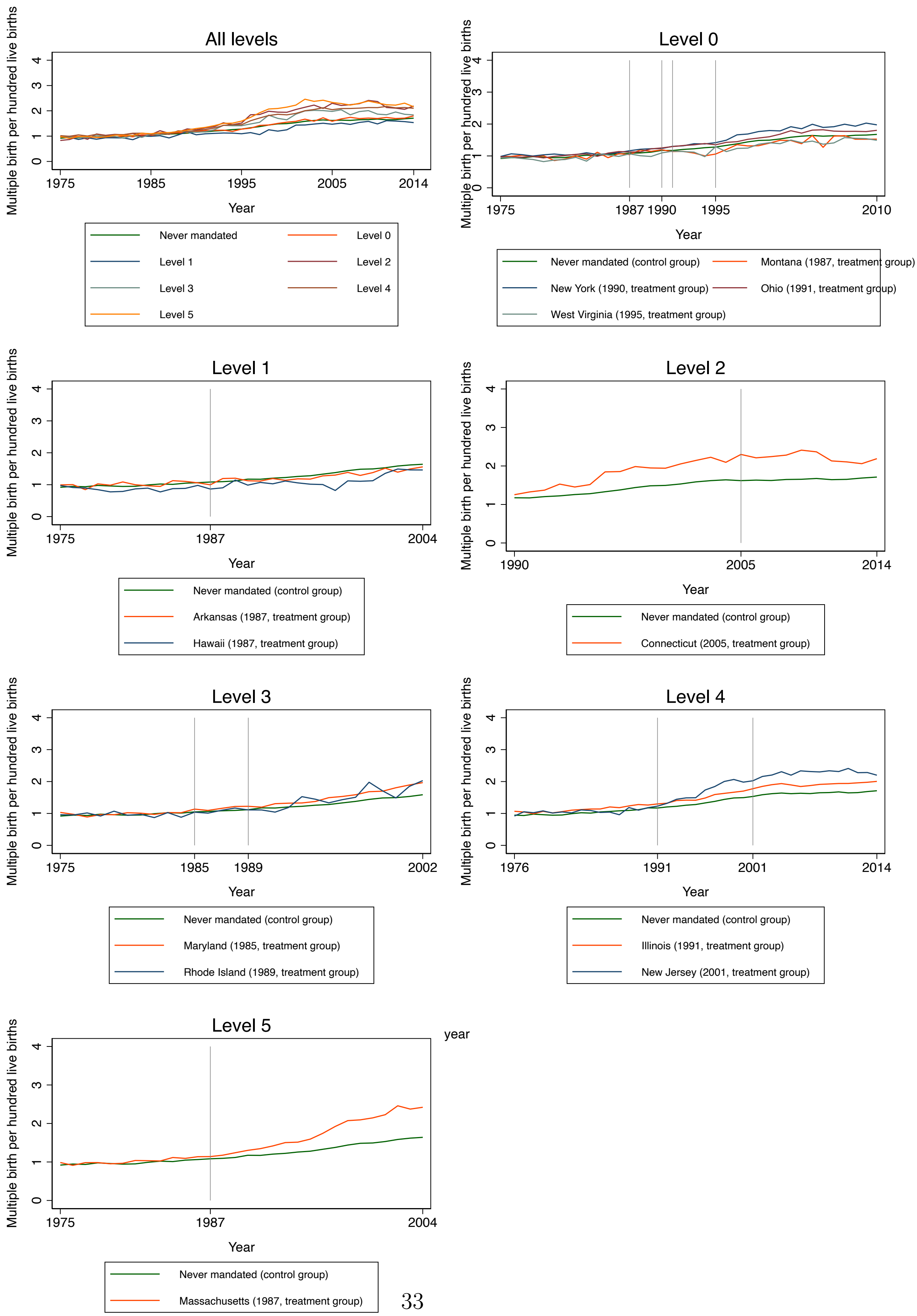
Figure 2: Multiple births per hundred live births by women's age



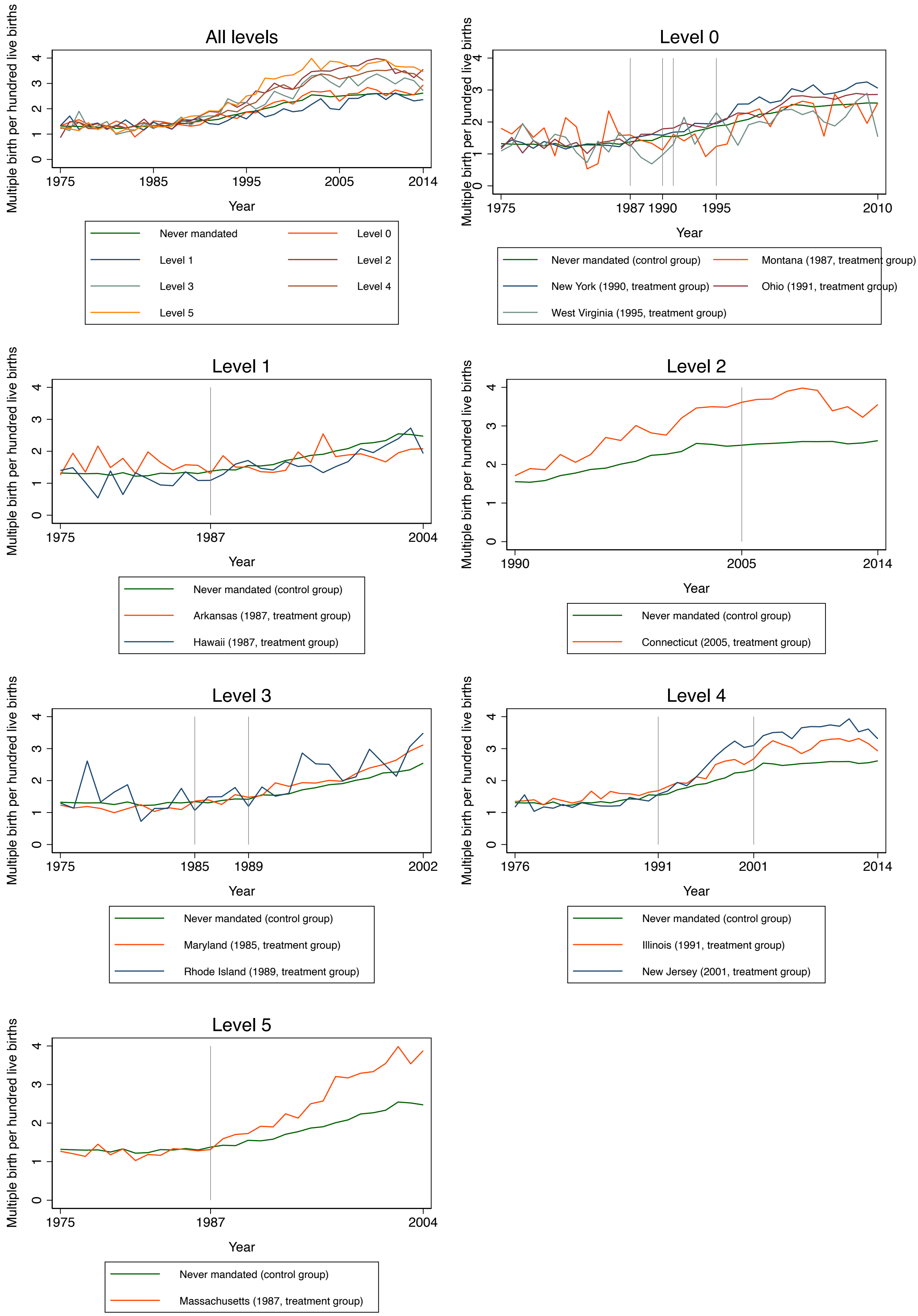
Note: Authors' calculations from the Detail Natality Data. Multiple births are defined as births that are not singleton.

Figure 3: Multiple births per hundred live births by IVF coverage generosity level and age of mother

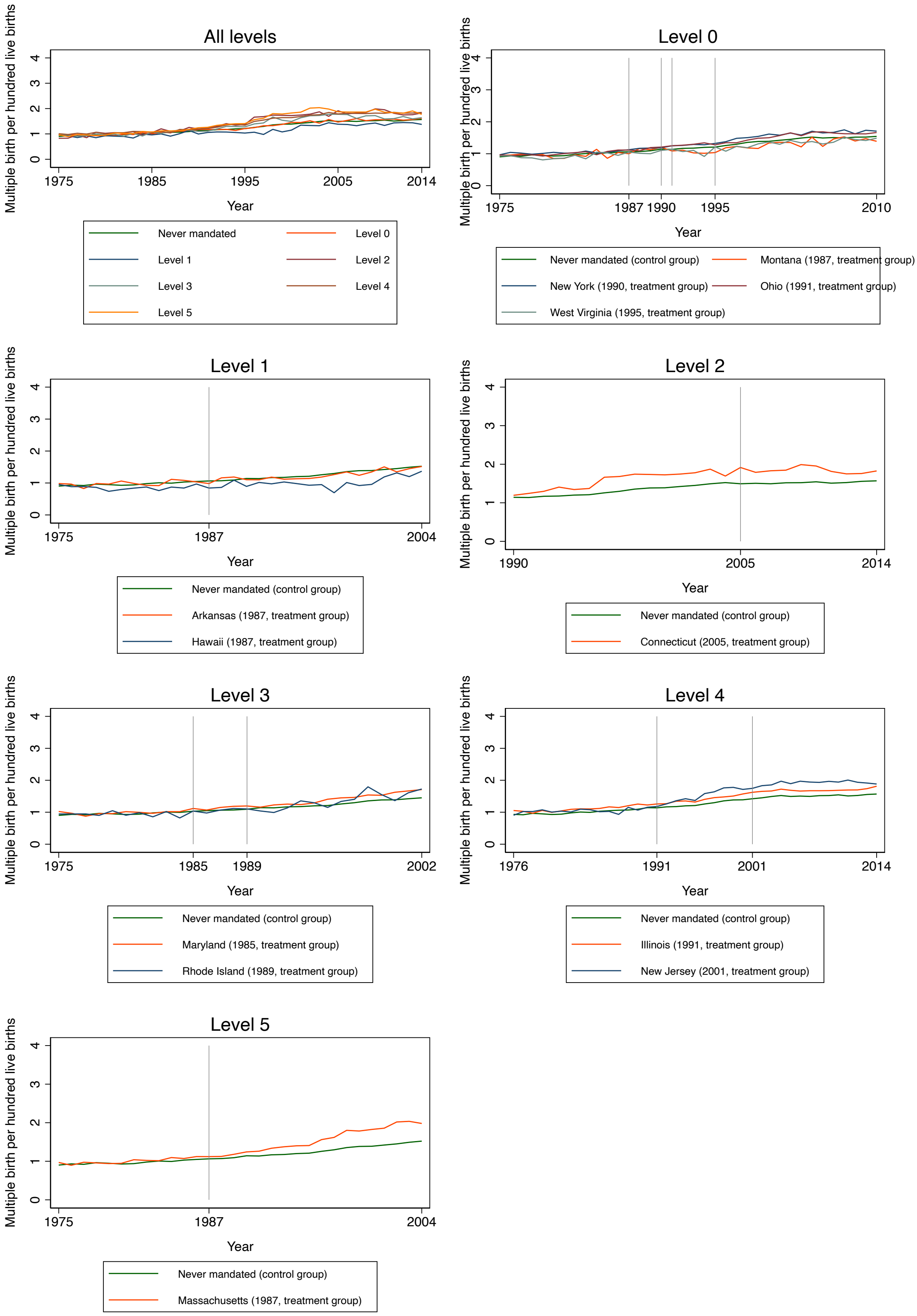
(a) All women



(b) Women 35 and older



(c) Women under 35

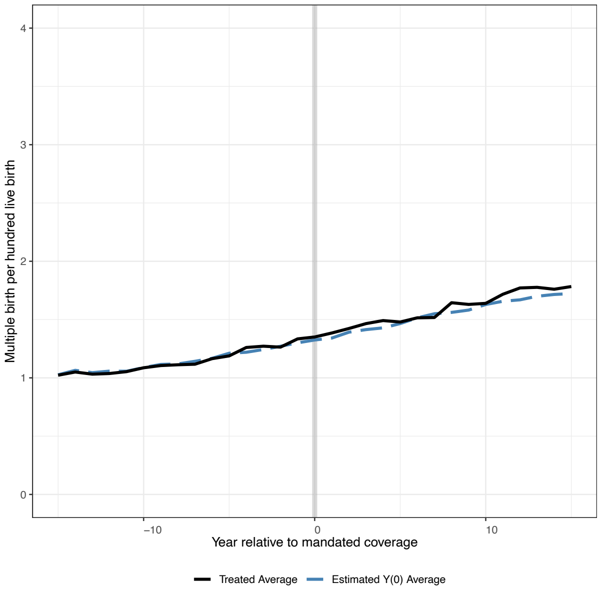


Note: The sample includes all births from National Vital Statistics Detail Natality Data from 1975–2014. Multiple births are defined as births that are not singletons.

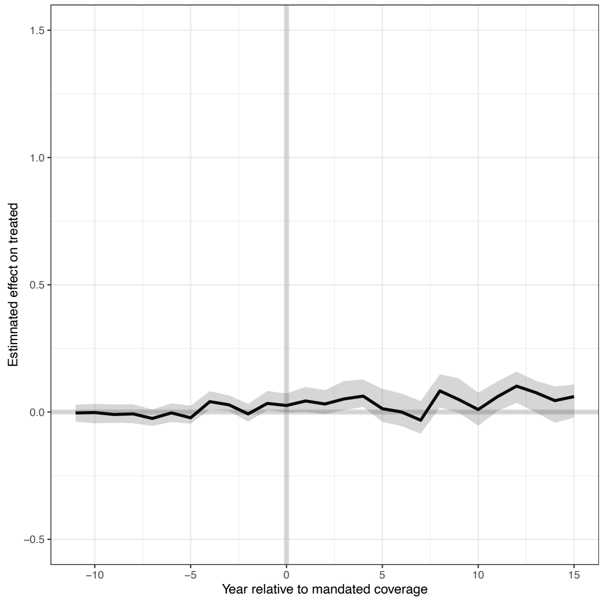
Figure 4: Effects of IVF coverage generosity level on multiple births per hundred live births, GSC model for all women

(a) All levels

(1) Treated average and estimated average for treated states

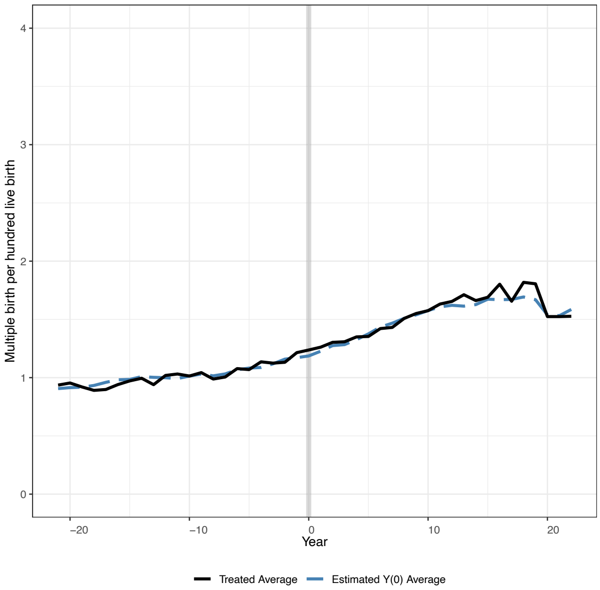


(2) Estimated treatment effect on treated

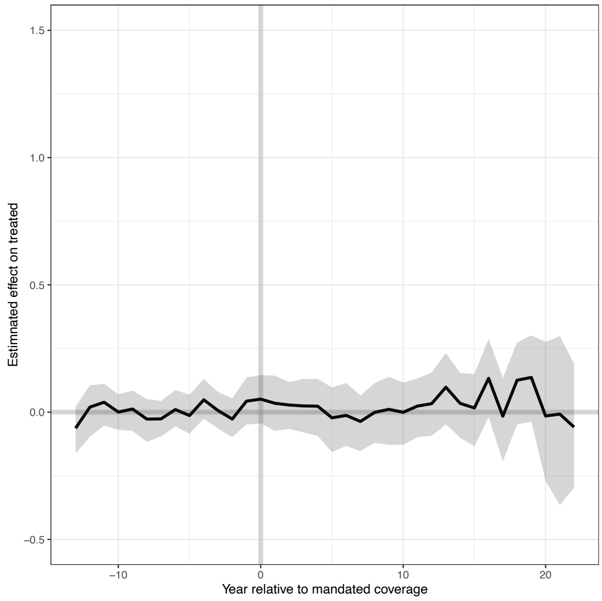


(b) Level 0

(1) Treated average and estimated average for treated states

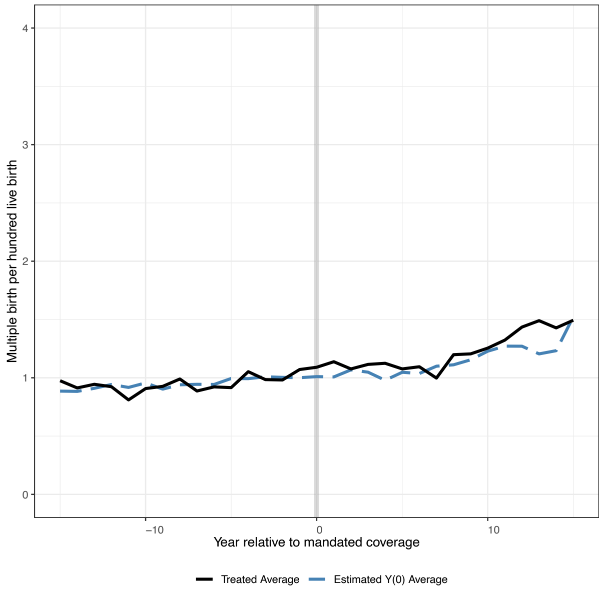


(2) Estimated treatment effect on treated

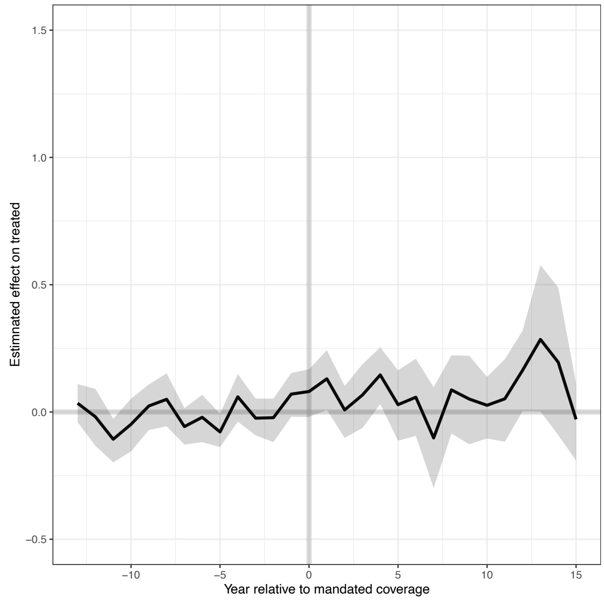


(c) Level 1

(1) Treated average and estimated average for treated states

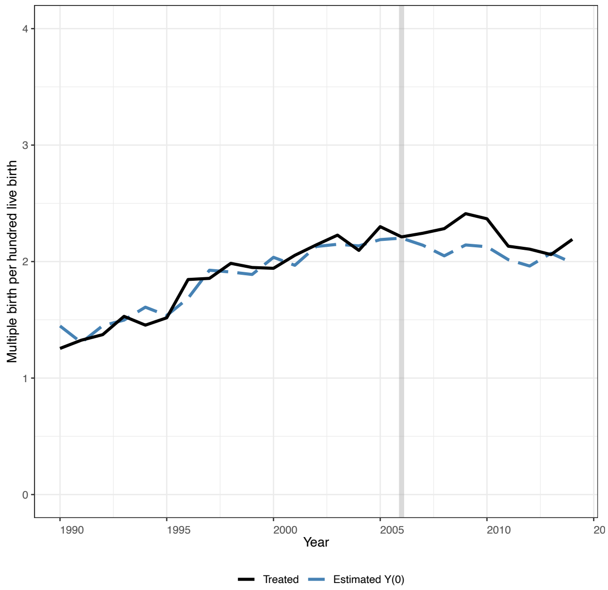


(2) Estimated treatment effect on treated

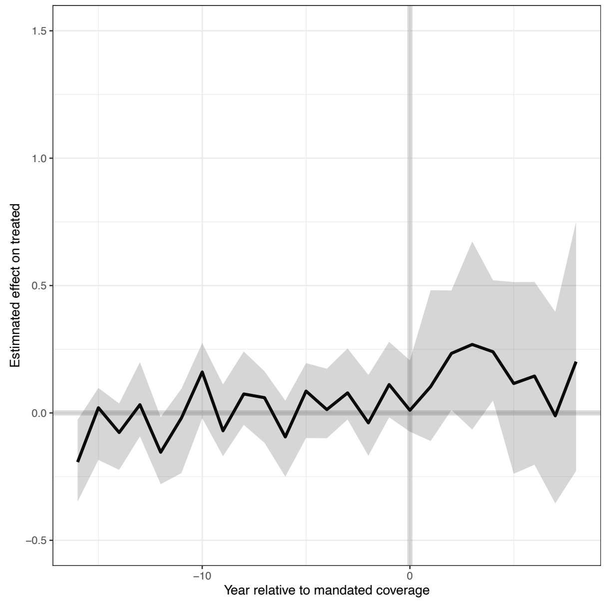


(d) Level 2

(1) Treated average and estimated average for treated states

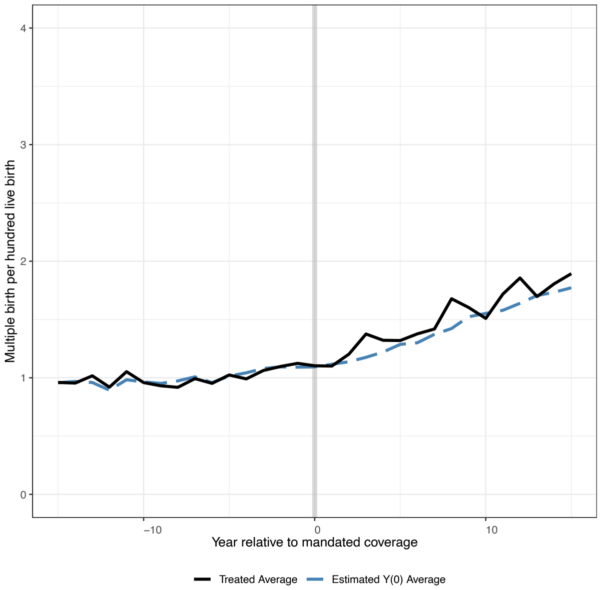


(2) Estimated treatment effect on treated

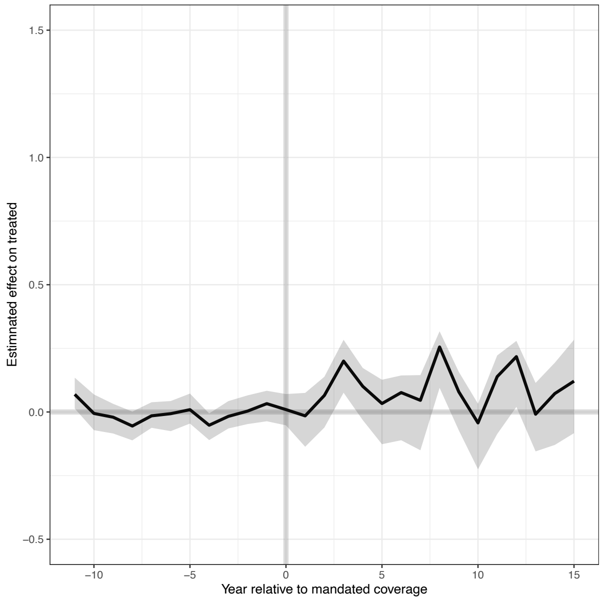


(e) Level 3

(1) Treated average and estimated average for treated states

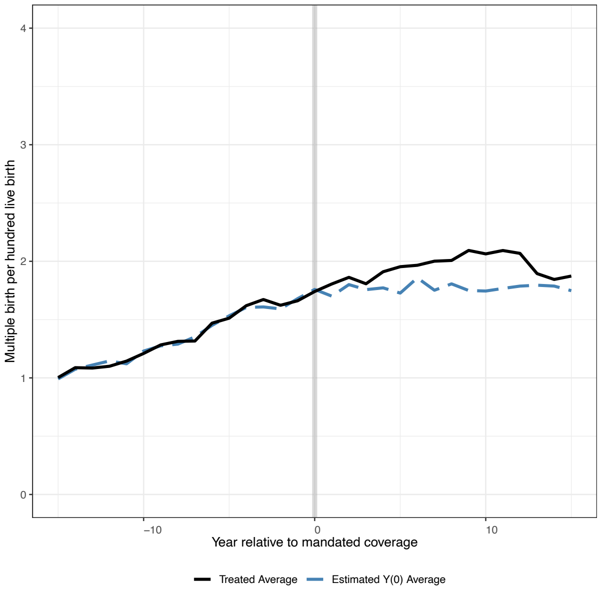


(2) Estimated treatment effect on treated

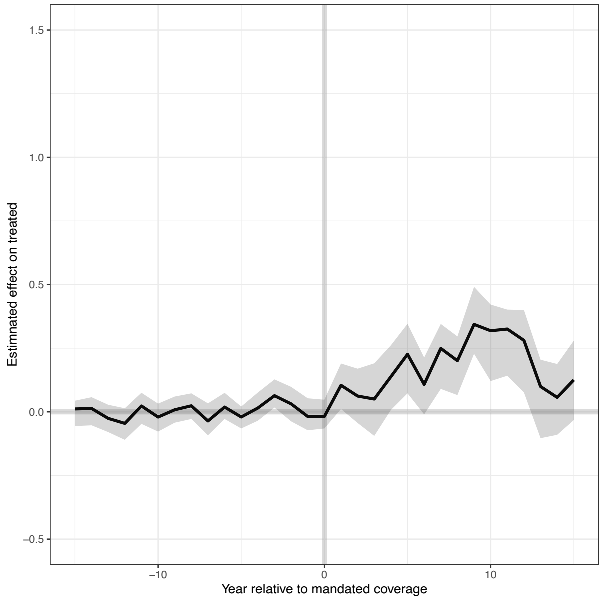


(f) Level 4

(1) Treated average and estimated average for treated states

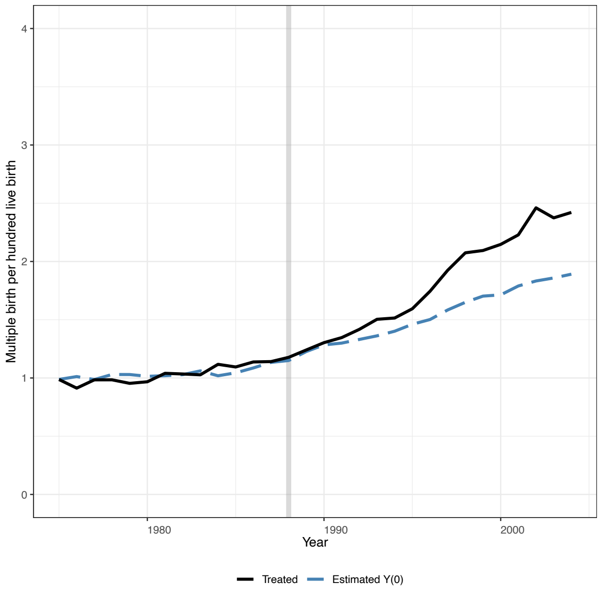


(2) Estimated treatment effect on treated

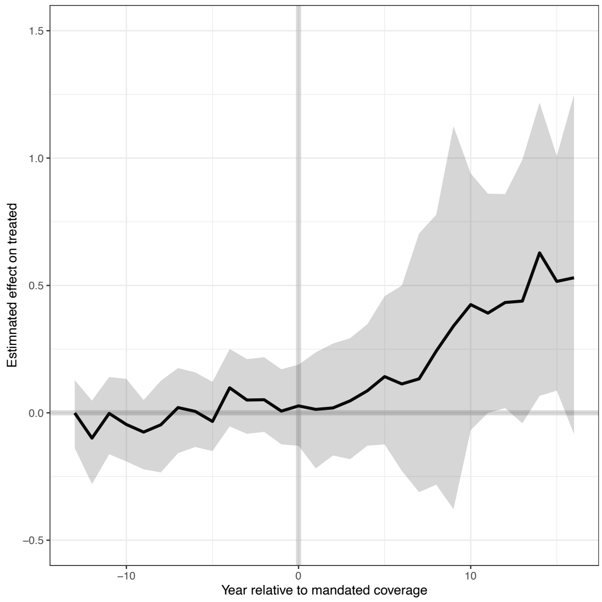


(g) Level 5

(1) Treated average and estimated average for treated states



(2) Estimated treatment effect on treated

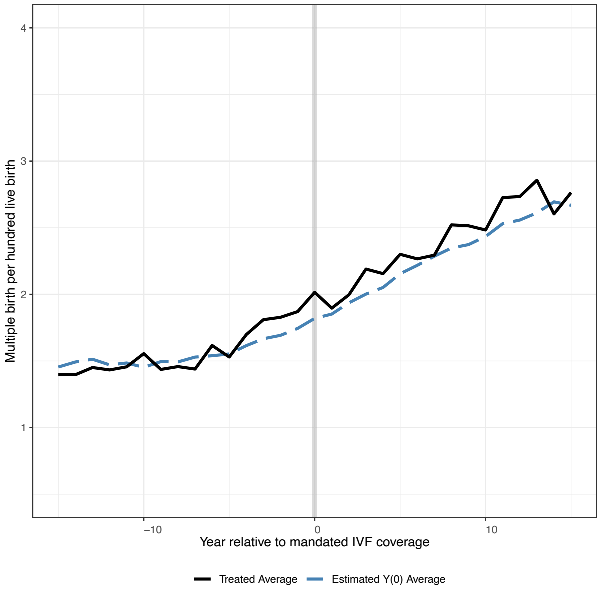


Note: This figure plots the estimated counter-factual outcome $Y(0)$ and the treatment effect on the treated on multiple births per hundred live births using the GSC model specified in Equation (2). The sample includes all births in the US from 1975-2014 from the National Vital Statistics, aggregated by state-year. The included covariates in the model are listed in the Notes to Table 5. The gray shade shows the %95 confidence intervals for the estimated effects.

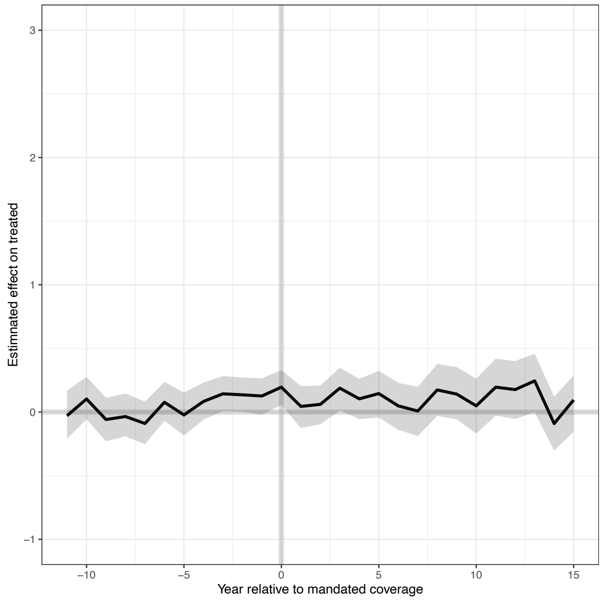
Figure 5: Effects of IVF coverage generosity level on multiple births per hundred live births, GSC model, women 35 and older

(a) All levels

(1) Treated average and estimated average for treated states

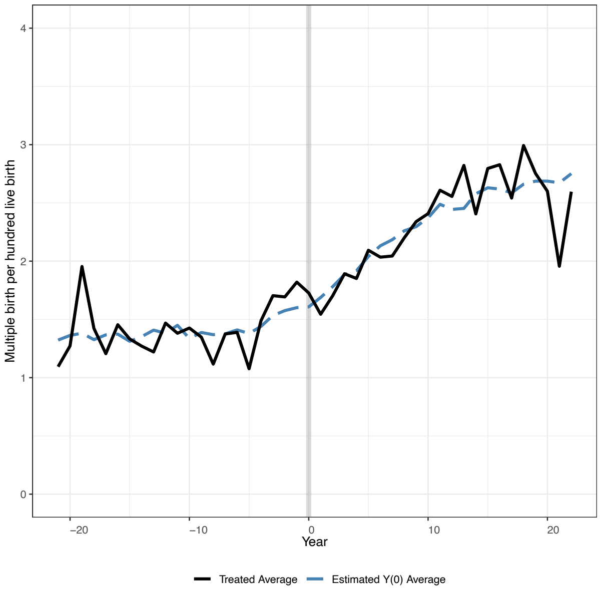


(2) Estimated treatment effect on treated

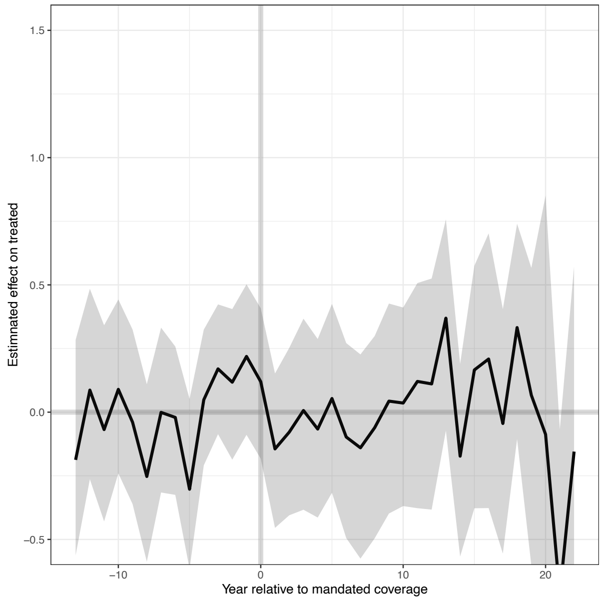


(b) Level 0

(1) Treated average and estimated average for treated states

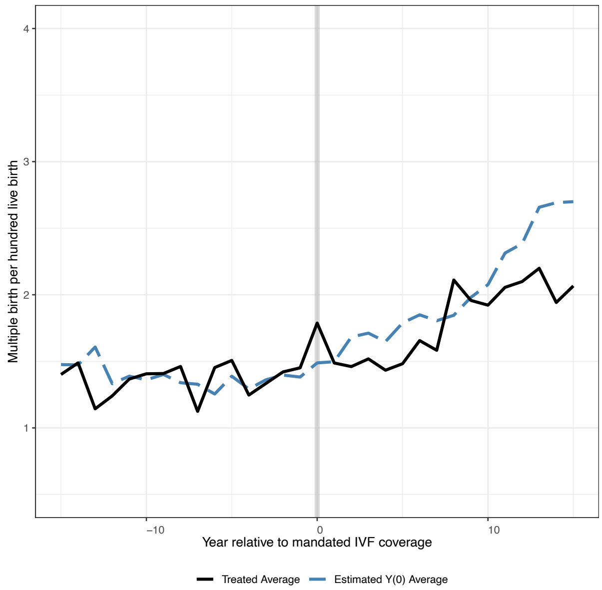


(2) Estimated treatment effect on treated

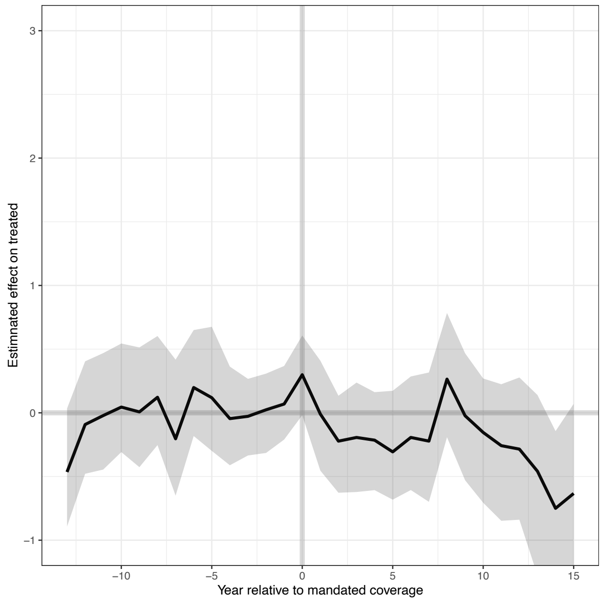


(c) Level 1

(1) Treated average and estimated average for treated states

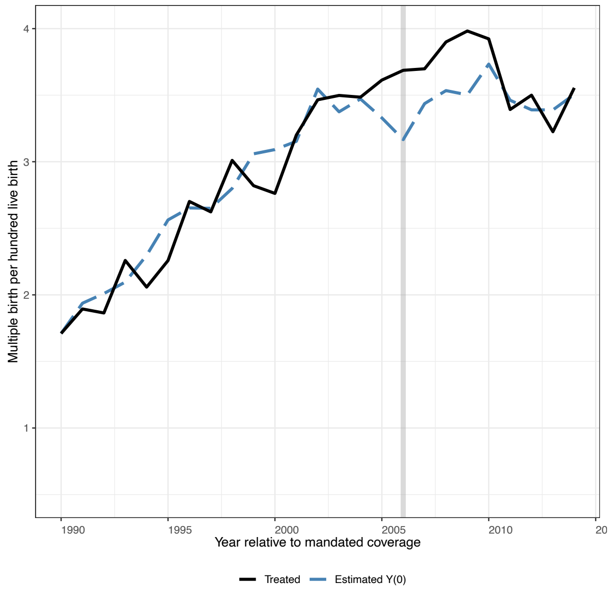


(2) Estimated treatment effect on treated

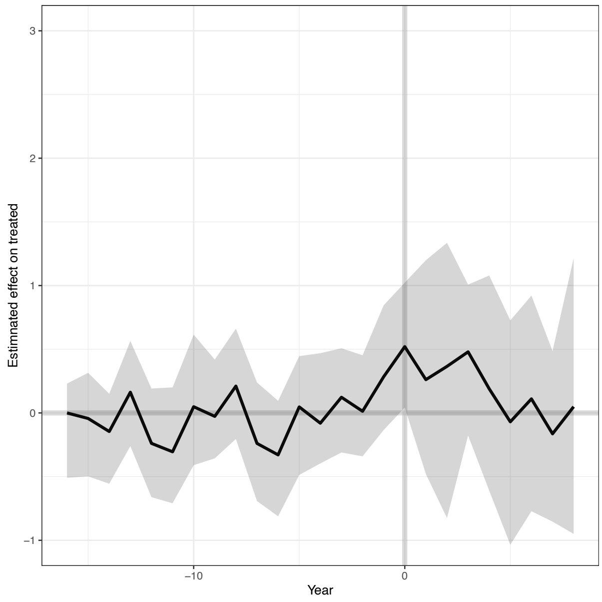


(d) Level 2

(1) Treated average and estimated average for treated states

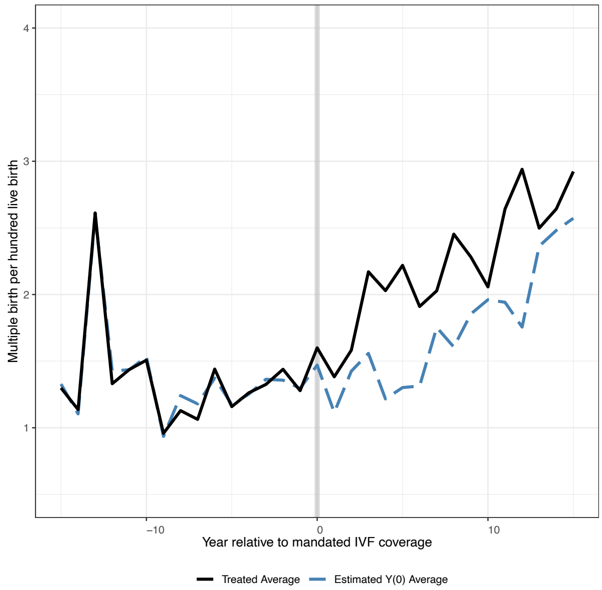


(2) Estimated treatment effect on treated

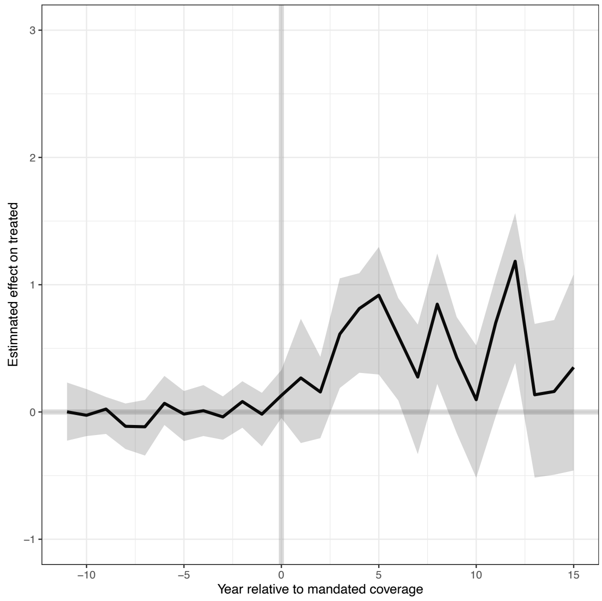


(e) Level 3

(1) Treated average and estimated average for treated states

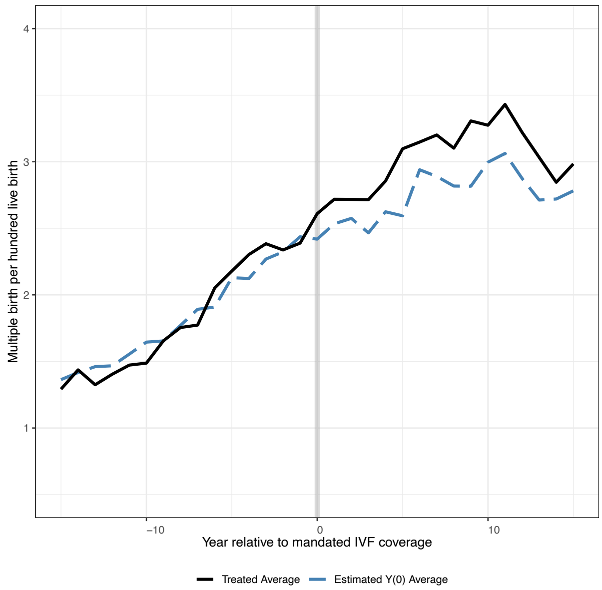


(2) Estimated treatment effect on treated

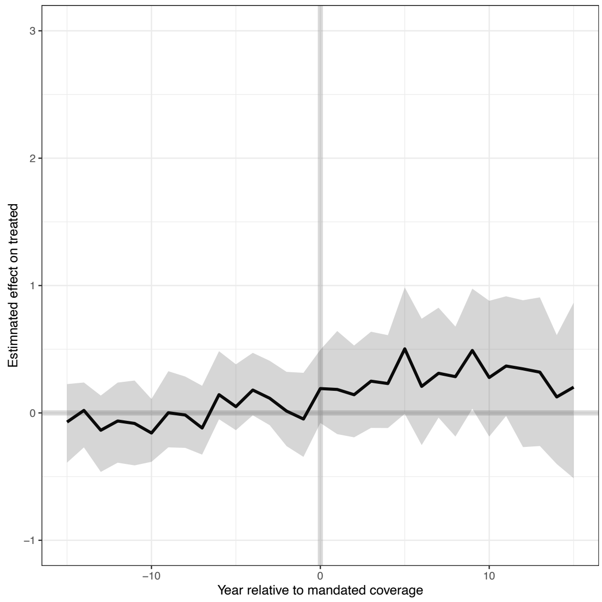


(f) Level 4

(1) Treated average and estimated average for treated states



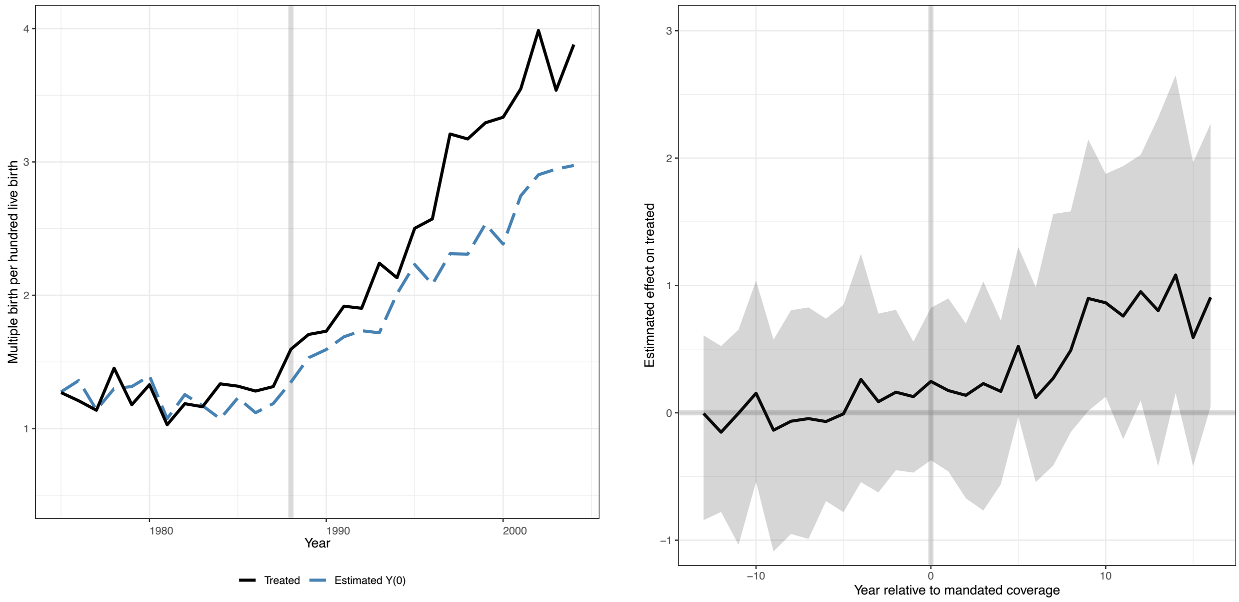
(2) Estimated treatment effect on treated



(g) Level 5

(1) Treated average and estimated average for treated states

(2) Estimated treatment effect on treated

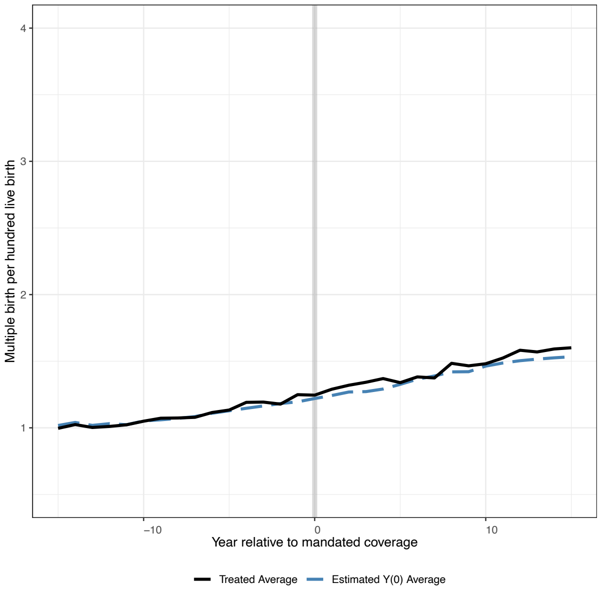


Notes: See notes for Figure 4.

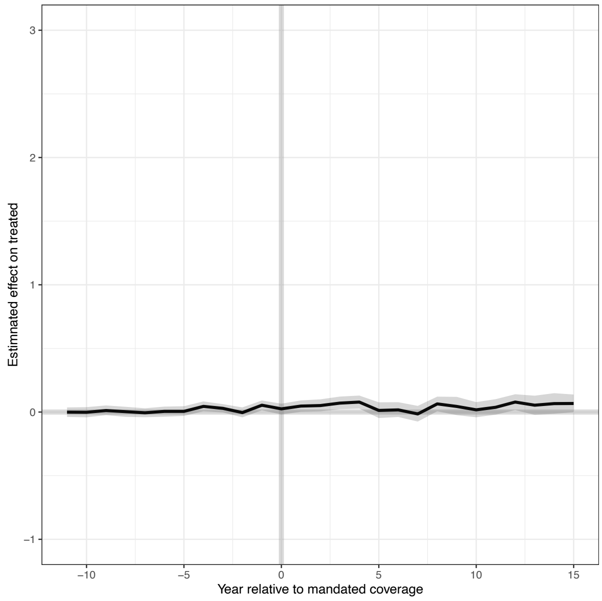
Figure 6: Effects of IVF coverage generosity on multiple births per hundred live births, GSC model, women under 35

(a) All levels

(1) Treated average and estimated average for treated states

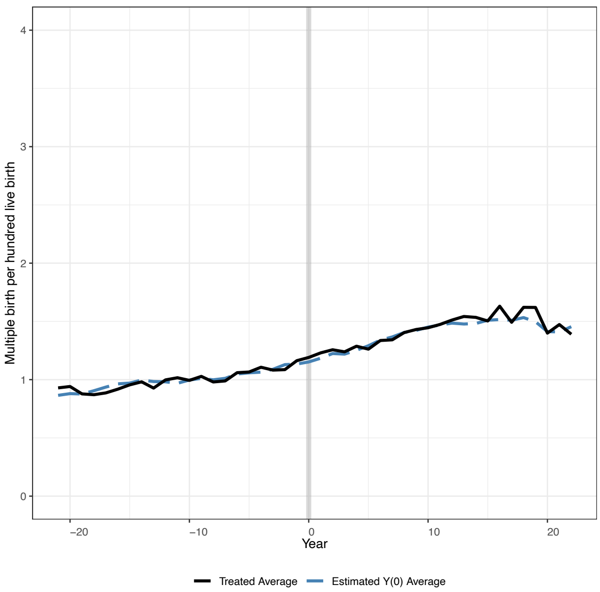


(2) Estimated treatment effect on treated

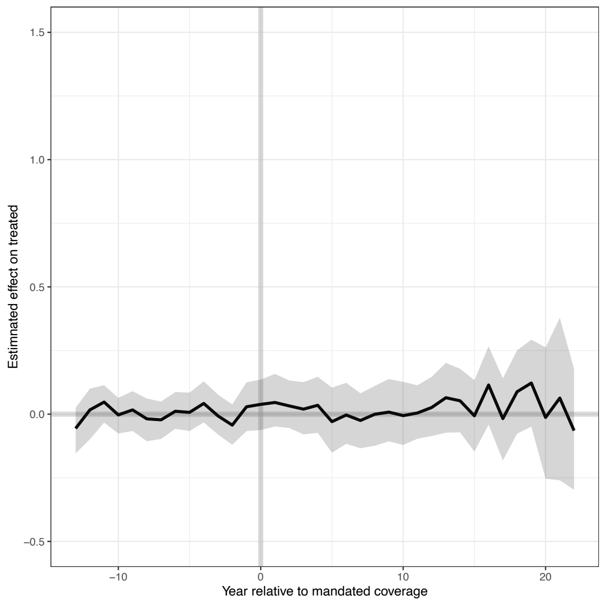


(b) Level 0

(1) Treated average and estimated average for treated states

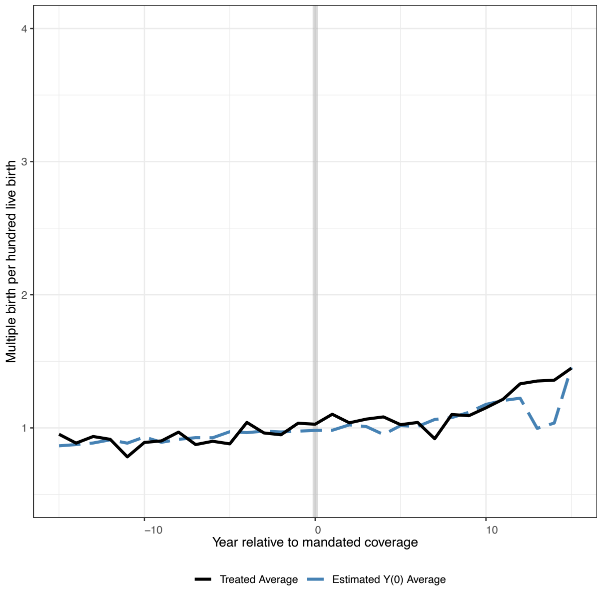


(2) Estimated treatment effect on treated

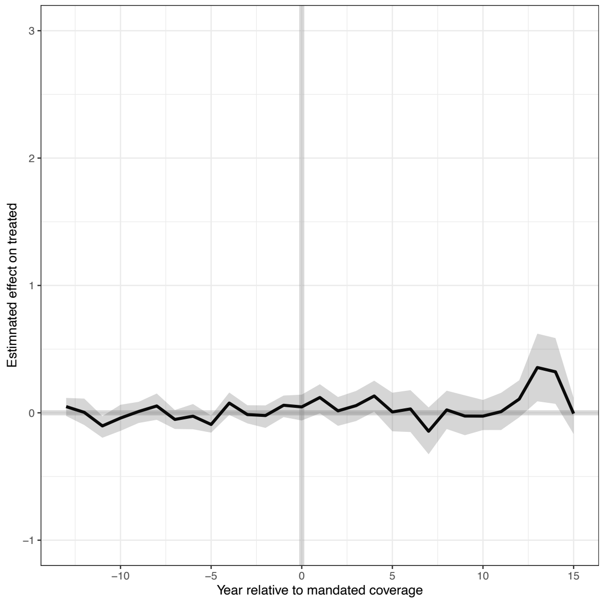


(c) Level 1

(1) Treated average and estimated average for treated states

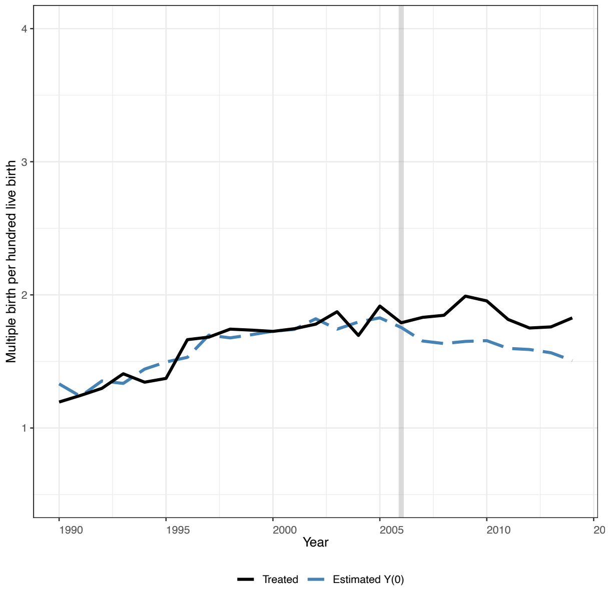


(2) Estimated treatment effect on treated

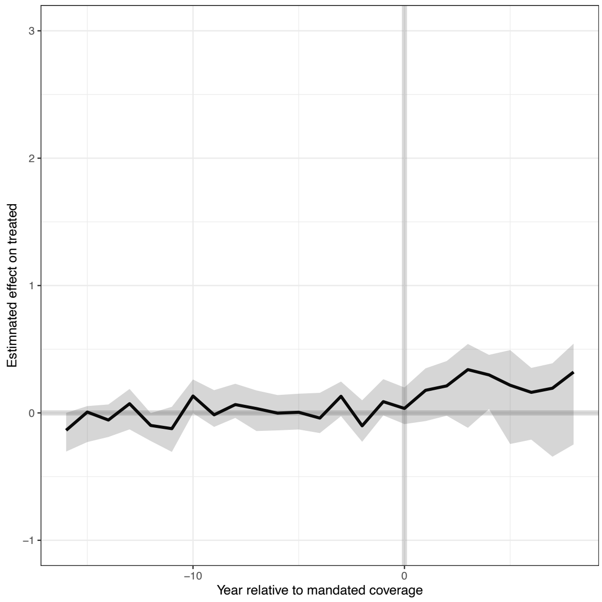


(d) Level 2

(1) Treated average and estimated average for treated states

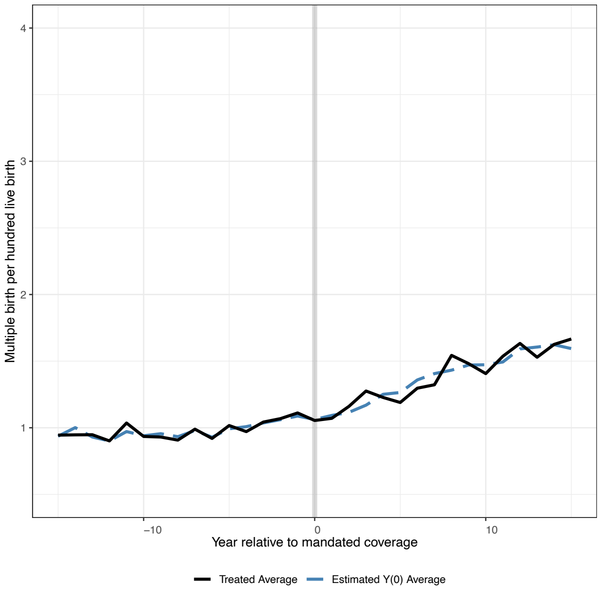


(2) Estimated treatment effect on treated

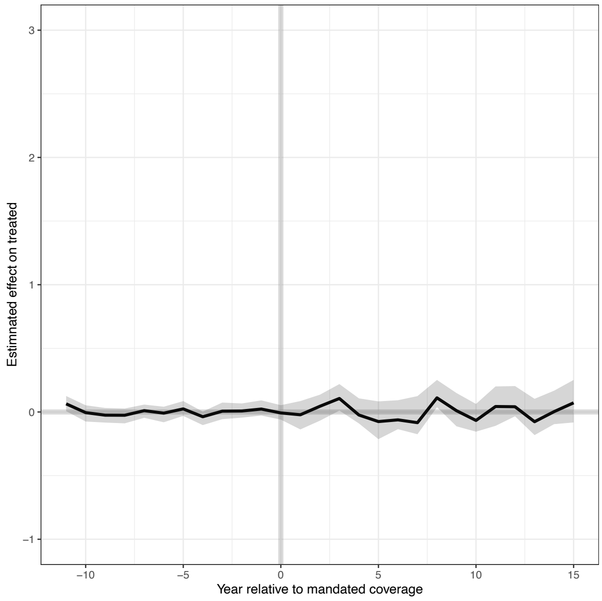


(e) Level 3

(1) Treated average and estimated average for treated states

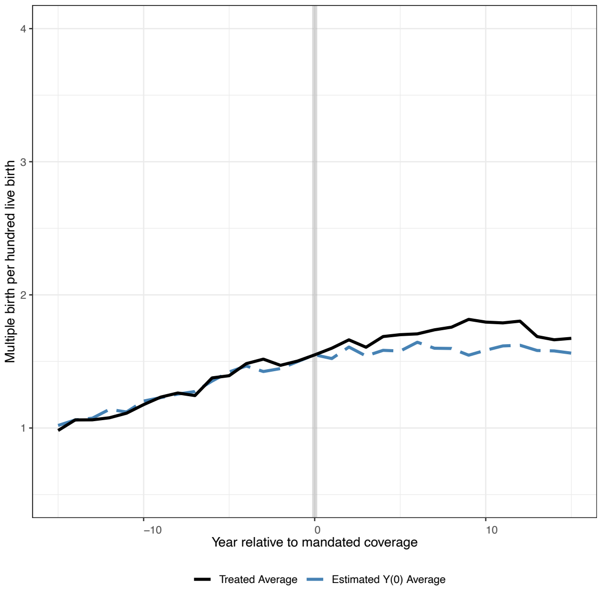


(2) Estimated treatment effect on treated

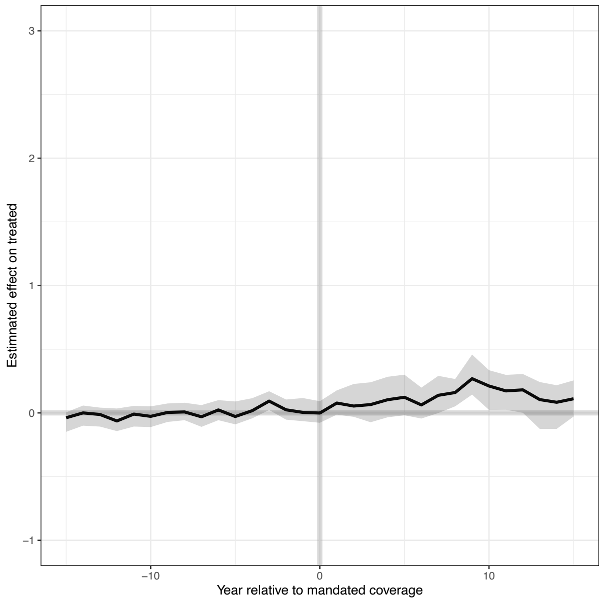


(f) Level 4

(1) Treated average and estimated average for treated states



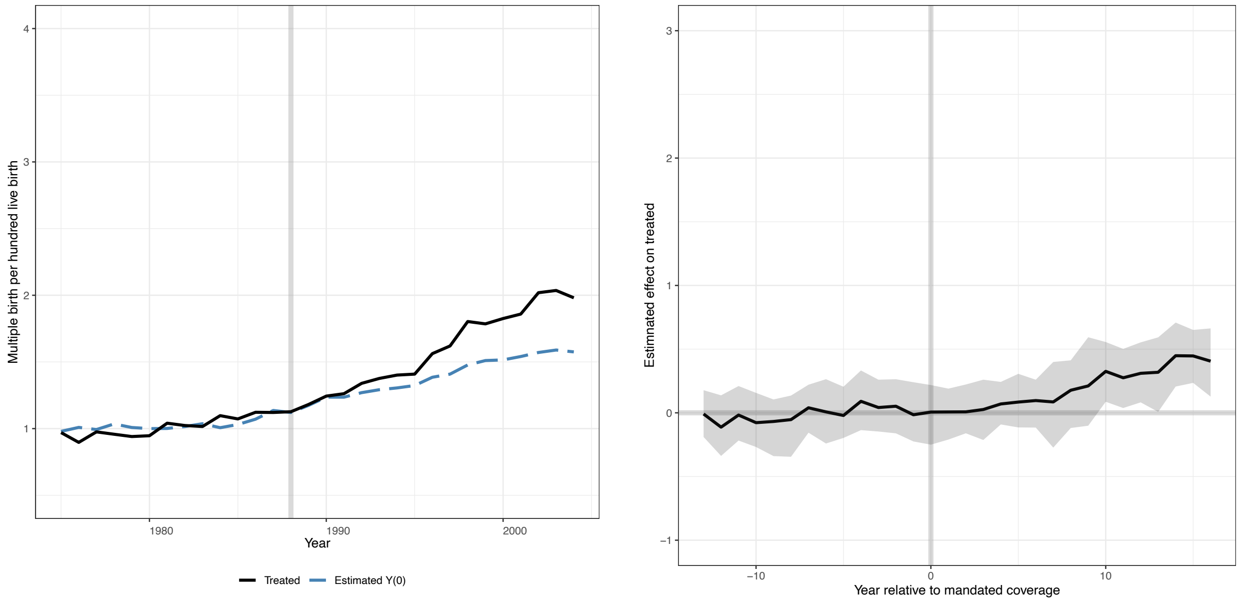
(2) Estimated treatment effect on treated



(g) Level 5

(1) Treated average and estimated average for treated states

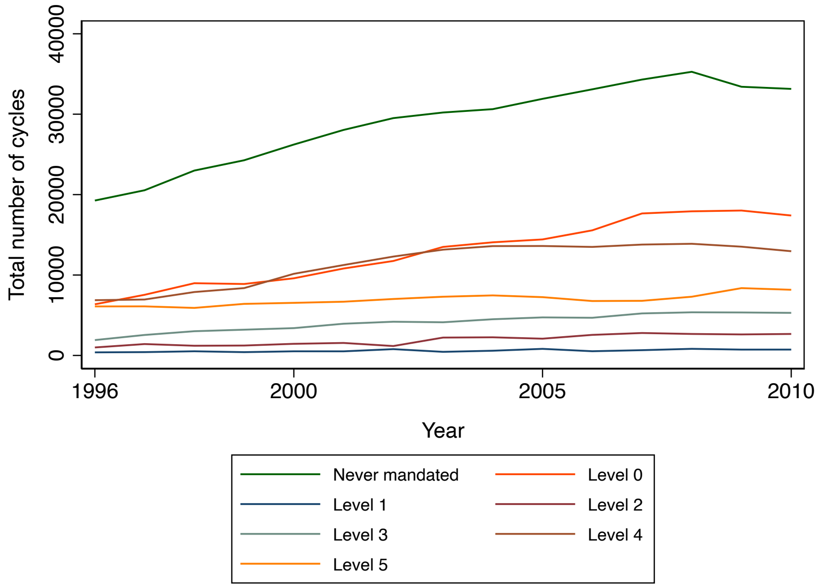
(2) Estimated treatment effect on treated



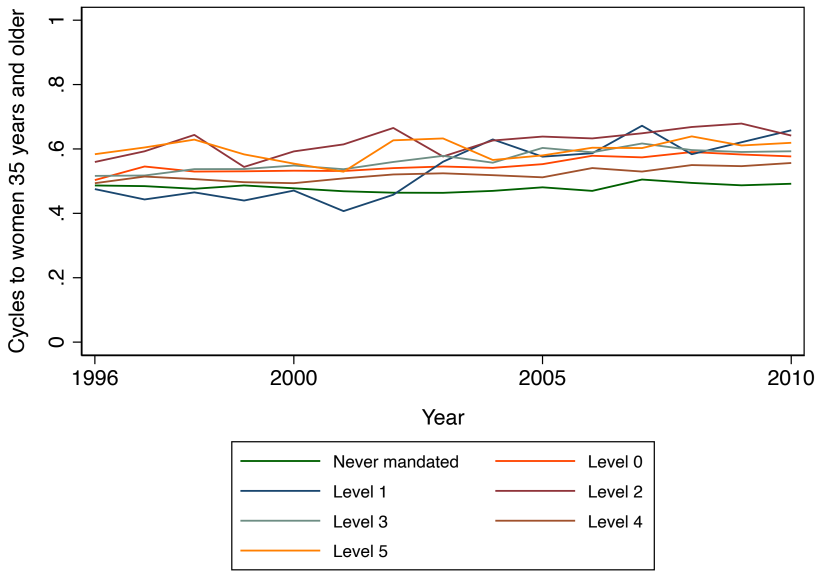
Notes: See notes to Figure 4.

Figure 7: Patients' IVF utilization behavior, by IVF coverage generosity level

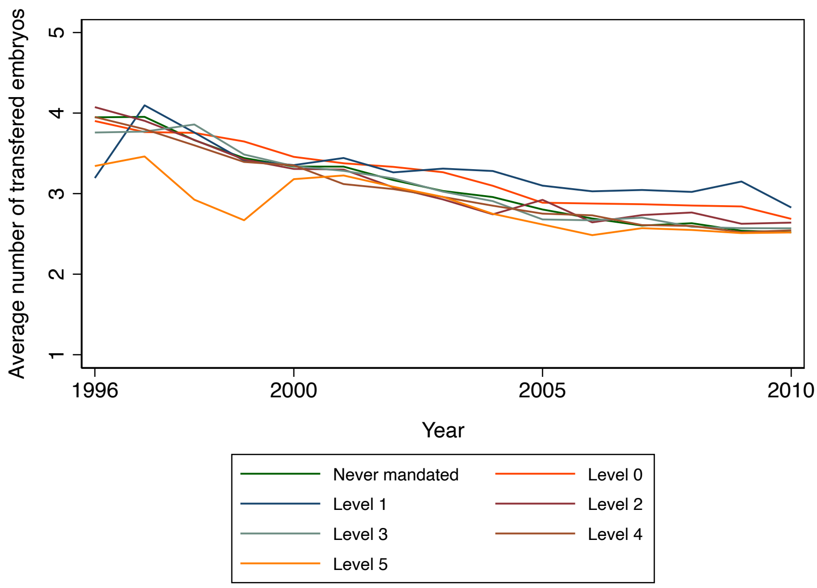
(a) Total number of cycles



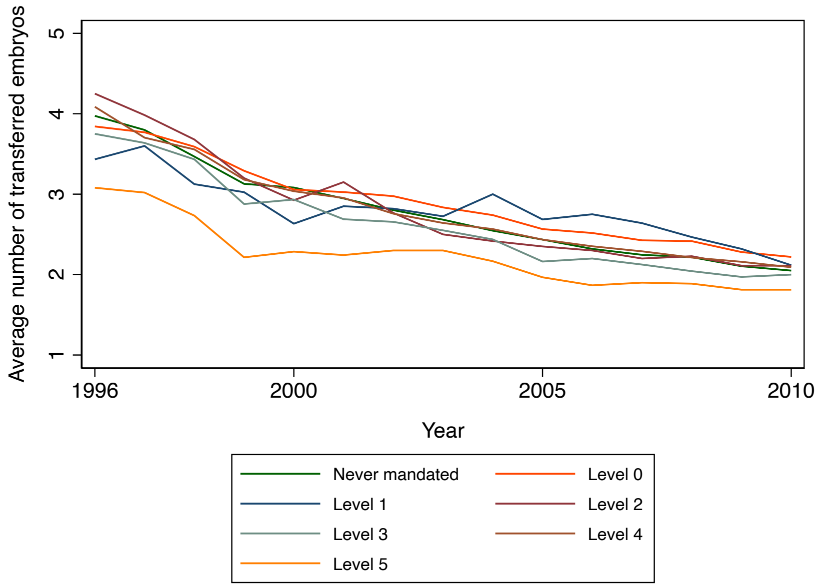
(b) Share of cycles to women 35 and older



(c) Average number of transferred embryos for women 35 and older



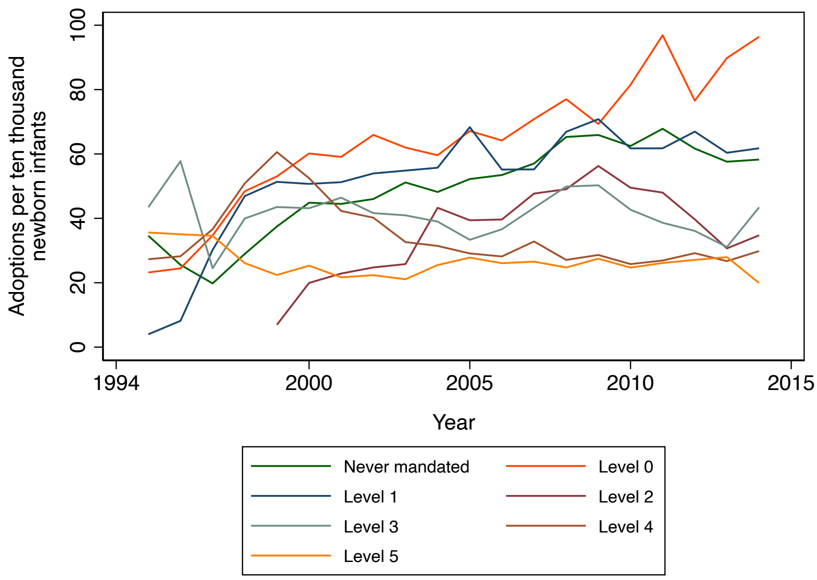
(d) Average number of transferred embryos for women under 35



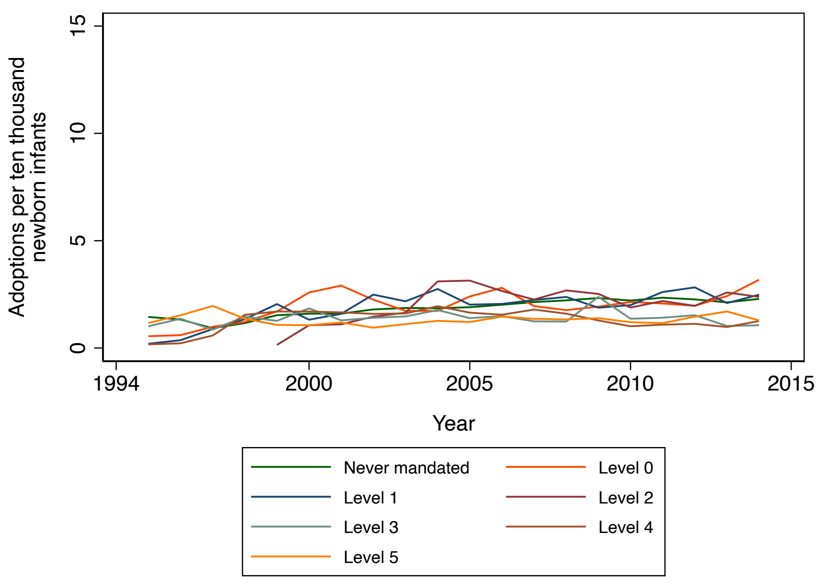
Note: This figure plots trends in patients' utilization behavior using SART's clinic-level data from 1996–2010.

Figure 8: Child adoption rates by IVF coverage generosity level and age of mother

(a) Women 35 and older



(b) Women under 35



Note: The study sample includes all adoptions of children ages 0–6 finalized between 1994–2015 from the National Data Archive on Child Abuse and Neglect (NDACAN). The denominator (total number of births) is from the National Vital Statistics Detail Natality Data.

Appendix

A Estimation procedure of a GSC model

There are two main approaches to estimating causal effects when the common trend assumption is likely to be violated. The first approach uses a matching method to condition on pre-treatment observable characteristics (Abadie, 2005; Abadie et al., 2010, 2015). This approach helps to balance the effects of time-varying confounders between the treatment and control groups. The second approach is to explicitly model the unobserved time-varying confounders by using an interactive fixed effect model which includes state-specific intercepts interacted 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 special case. Xu (2017) provides a procedure for estimating a Generalized Synthetic Control (GSC) model specified in Equation (2) as:

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

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 (A.1)). 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. r is the number of factors (f_t and λ_i are r vectors). 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 and $|T|$ is the total number of time periods in the analysis. 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:

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

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

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

We define the prediction error as $e_{is} = y_{is}(0) - \widehat{y}_{is}(0)$. The mean square prediction error (MSPE) for given r is defined as:

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

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:

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

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 counter-factual based on $\widehat{\beta}$, $\widehat{F}$ and $\widehat{\lambda}_i$. That is:

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

The estimated average treatment effect on the treated at time t, ATT_t then is:

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

[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).

B DD and DDD estimates

Table B.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)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covars	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Number of cells	1,768	1,768	1,768	1,768	1,768	1,768	1,768	1,768	1,768	1,768	1,768	1,768	3,616	3,276	3,616	3,276

Note: Study sample includes all births in the US from 1975-2014. Data aggregated into state-year cells for DD analysis and state-year-age cell for DDD analysis. All models include state- and year-fixed effects. Included covariates listed in notes for Table 5. Standard errors are clustered in state level and appear in parentheses.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table B.2: Effects of IVF coverage generosity level on the number of infants per thousand births, DD and DDD models

	Difference-in-Differences												Difference-in-Difference-in-Differences			
	All women				Women 35 and older				Women under 35							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
All levels	1.07*	0.71***			1.72	1.22			0.59	0.57***			5.35***	2.75***		
	(0.55)	(0.26)			(1.23)	(0.89)			(0.39)	(0.20)			(0.95)	(0.93)		
Level 0			0.14	0.05			-0.45	-0.37			0.09	0.06			4.99***	2.57**
			(0.64)	(0.40)			(1.61)	(1.15)			(0.43)	(0.30)			(1.30)	(1.22)
Level 1			-1.26***	-0.20			-3.33**	-3.26***			-1.15***	-0.07			2.50	-1.22
			(0.18)	(0.30)			(1.58)	(1.19)			(0.39)	(0.23)			(2.22)	(1.33)
Level 2			1.38***	1.77***			3.79***	3.25***			0.33***	1.11***			6.29***	3.91***
			(0.12)	(0.27)			(0.31)	(0.66)			(0.10)	(0.22)			(0.00)	(0.38)
Level 3			2.23***	0.72			3.90***	3.30***			1.56***	0.65			6.31***	4.19***
			(0.36)	(0.54)			(0.37)	(1.12)			(0.25)	(0.42)			(0.59)	(0.75)
Level 4			2.33**	2.09***			4.60***	3.44***			1.38**	1.70***			8.19***	5.63***
			(0.90)	(0.74)			(1.37)	(0.92)			(0.54)	(0.50)			(0.31)	(0.32)
Level 5			4.60***	2.18***			9.03***	7.10***			2.96***	1.36***			10.28***	7.95***
			(0.17)	(0.47)			(0.38)	(1.01)			(0.15)	(0.38)			(0.00)	(0.44)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covars	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Number of cells	1,768	1,768	1,768	1,768	1,768	1,768	1,768	1,768	1,768	1,768	1,768	1,768	3,616	3,276	3,616	3,276

Note: See notes for Table B.1. Standard errors are clustered in state level and are presented in parentheses.
$*p < 0.10, **p < 0.05, ***p < 0.01$