

Machado, Matilde P.; Sanz-de-Galdeano, Anna

## Article

# Coverage of infertility treatment and fertility outcomes

SERIEs - Journal of the Spanish Economic Association

## Provided in Cooperation with:

Spanish Economic Association

*Suggested Citation:* Machado, Matilde P.; Sanz-de-Galdeano, Anna (2015) : Coverage of infertility treatment and fertility outcomes, SERIEs - Journal of the Spanish Economic Association, ISSN 1869-4195, Springer, Heidelberg, Vol. 6, Iss. 4, pp. 407-439, <https://doi.org/10.1007/s13209-015-0135-0>

This Version is available at:

<https://hdl.handle.net/10419/158546>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>

# Coverage of infertility treatment and fertility outcomes

Matilde P. Machado<sup>1</sup> · Anna Sanz-de-Galdeano<sup>2</sup>

Received: 9 August 2015 / Accepted: 6 November 2015 / Published online: 25 November 2015  
© The Author(s) 2015. This article is published with open access at Springerlink.com

**Abstract** Policy interventions that increase insurance coverage for infertility treatments may affect fertility trends, and ultimately, population age structures. However, such policies have ignored the overall impact of coverage on fertility. We examine short-term and long-term effects of increased insurance coverage for infertility on the timing of first births and on women's total fertility rates. Our main contribution is to show that infertility mandates enacted in the United States during the 80s and 90s did not increase the total fertility rates of women by the end of their reproductive lives. We also show evidence that these mandates induced women to put off motherhood.

**Keywords** Assisted reproductive technologies · Infertility insurance mandates · Completed fertility · Delay of motherhood · Synthetic control method

**JEL Classification** I18 · J13

## 1 Introduction

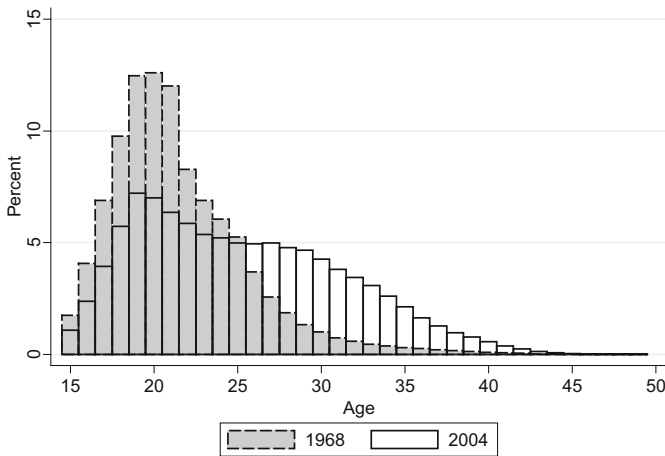
The average age at first birth in the United States has been rising steadily over the past decades, from 21.49 in 1968 to 23.72 in 1985 and 25.26 in 2004. As shown in Fig. 1, this increase has been accompanied by remarkable changes in the age distribution of

---

✉ Matilde P. Machado  
matilde.machado@uc3m.es  
Anna Sanz-de-Galdeano  
anna.sdg@ua.es

<sup>1</sup> Departamento de Economía, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe, Spain

<sup>2</sup> Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, Alicante, Spain



Source: Natality Data, National Center for Health Statistics. Authors' calculations.

**Fig. 1** Distributions of maternal age at first birth in 1968 and 2004

first-time mothers, which has become less skewed with a substantially higher density of first-time mothers older than 25 and an extension of first-time motherhood beyond the age of 40.

Women, however, face a biological time constraint on bearing children because fecundity decreases with age. The introduction of and the subsequent increase in the use of assisted reproductive therapies (ARTs) have helped women in extending their reproductive lives (CDC 2007). ARTs, particularly in-vitro fertilisation (IVF), are very expensive procedures. For example, in 1992, a birth from an IVF procedure cost between 44,000 and 211,942 USD (Neumann et al. 1994). Over time, however, ART patients have faced substantially lower costs due to increased competition (Hamilton and McManus 2012), a reduced number of cycles due to better technology,<sup>1,2</sup> and most importantly, the availability of insurance in both the United States and in Europe.<sup>3</sup> In this paper, we analyse whether easier access to ARTs induces women to delay motherhood and whether, in the long term, it affects women's completed fertility by the end of their reproductive lives.

The perception that ARTs increase fertility has led the European Parliament to call on member states to insure 'the right to universal access to infertility treatment' (Ziebe and Devroey 2008). This movement's incarnation in the United States has sponsored several attempts at approving the 'Family Building Act of 2009', which would extend coverage for infertility treatments, and the enactment by several states of infertility

<sup>1</sup> A cycle is the process that starts with the administration of fertility medication to stimulate a woman's ovaries to produce several follicles. Fertilization may occur in the laboratory (IVF) or in the womb.

<sup>2</sup> See, for example, the evolution of success rates in the 2005 CDC assisted reproductive technology (ART) report at <http://www.cdc.gov/art/PDF/508PDF/2005ART508.pdf>.

<sup>3</sup> In Europe, some countries such as Belgium, Denmark, France, Greece, Israel, Slovenia, and Sweden have complete public coverage for infertility treatment (IFFS 2007). The case of the United States is examined in this paper.

insurance coverage laws, which are referred to as infertility treatment mandates.<sup>4</sup> Considering the high cost of infertility treatments (Bitler and Schmidt 2012; Collins 2001), policy interventions that grant insurance coverage for infertility treatments may affect fertility trends and ultimately, population age structures. The mid- to long-term consequences of ARTs are central to the European debate on possible solutions to an ageing population—i.e., can ARTs be part of a package of policies intended to increase fertility rates in Europe? (Grant et al. 2006; Ziebe and Devroey 2008).<sup>5</sup>

The answer is complex because the short-term effect of an increase in coverage for infertility treatments may be very different from the long-term effect. In the short term, an increase in the aggregate fertility rate is expected due to an increase in fertility amongst the least fertile women (a compositional effect). Typically, these are relatively old women who delayed motherhood and would be unlikely to conceive otherwise (Buckles 2005; Schmidt 2005a, 2007). Moreover, increased access to ARTs increases the frequency of multiple births in the population (Bundorf et al. 2007). These two effects are short-term and non-strategic and may be referred to as ex-post moral hazard. In the long term, however, easier access to infertility treatments and the possibility of extending reproductive life may induce women to further delay motherhood, possibly because of overly optimistic perceptions about the effectiveness of infertility treatments (Lampi 2006; Benyamini 2003). This response by relatively young women, which may be referred to as ex-ante moral hazard, is strategic and would increase the average age at first birth for several years after the policy was implemented.<sup>6</sup> Such a response would be consistent, for example, with the delay in marriage due to increased infertility coverage documented in Abramowitz (2014). Therefore, it is possible that an increase in insurance coverage for infertility treatment may have negative effects on total fertility in the mid- to long-term. This paper examines these issues in the United States, where, by 2001, more than 1 % of live births were due to IVF (CDC 2007).

Our objective in this paper is twofold. First, we analyse the impact of an increase in infertility insurance on the timing of first births. Although this question was first explored by Buckles (2005), we believe that our paper contributes in a substantial way to the few existing manuscripts that address this topic by using more adequate data and methodology. Moreover, we go a step further by looking into the long term effects of increasing infertility insurance. Second, we ask whether the increase in infertility insurance affects completed fertility, i.e., fertility by the end of a woman's reproductive

<sup>4</sup> Direct evidence of the impact of infertility coverage mandates on ART utilisation is provided by Bitler and Schmidt (2012), who show that the self-reported use of infertility treatments increases among highly educated older women, and by Mookim et al. (2008), who use the mandates as an instrument for ARTs usage using data on medical claims. Indirect evidence of the impact of infertility insurance on ART utilisation is provided, among others, by studies showing an increase in multiple births among white women (Bitler 2005; Bundorf et al. 2007) and a higher prevalence of unhealthy twins (Bitler 2005).

<sup>5</sup> The total fertility rate for the 25 countries in the European Union is only 1.5 births per woman (Ziebe and Devroey 2008).

<sup>6</sup> The terms 'ex-ante' and 'ex-post' moral hazard are common in the Health Economics literature; while the latter is typically associated with demand responses to price changes, the former is associated with changes in behaviour that affect the probability of disease or need for medical attention. In consequence, an increase in the usage of ART in response to a price reduction due to a larger availability of insurance may be referred to ex-post moral hazard, while changes in behavior such as postponement of motherhood, which increases the likelihood of usage of ARTs in the future, may be referred to as ex-ante moral hazard.

life. This study represents, to the best of our knowledge, the first to address this issue.<sup>7</sup> Both objectives are analysed using data from the United States.

To assess whether infertility insurance induces a delay in motherhood, one needs to combine evidence about reduced fertility of young women with information on when women become mothers, i.e., when (if at all) they stop delaying motherhood. This precisely describes the approach we adopt in the first part of this paper; we not only offer similar evidence as [Buckles \(2005\)](#) on the reduced probability that relatively young women in mandated states have children, but we also demonstrate that the average age of first-time mothers continues to increase in the medium to long term after the enactments of infertility mandates.<sup>8</sup> Our long-term estimate (10–16 years after the first and the last mandates were passed) ranges from 3 to 5 months. These effects are substantial insofar as they represent between 15.7 and 18.8 % of the total increase in the age of first-time mothers during the period considered for the group of six states that enacted infertility treatment mandates<sup>9</sup> and between 24.8 and 34.3 % for the three states with the most generous coverage (Illinois, Massachusetts, and Rhode-Island).<sup>10</sup>

The ageing of first-time mothers may impact women's completed fertility in the long term. Hence, our second goal is to determine whether infertility insurance indeed increases women's completed fertility by the end of their reproductive lives, a question that has not been addressed in the existing literature. In principle, any potential negative effects on fertility induced by a delay of motherhood may eventually be offset by a higher prevalence of multiple births,<sup>11</sup> so the impact of infertility insurance on completed fertility is ultimately an empirical question. Overall, our estimates, based on data on the number of biological children from the June CPS, show no statistically significant effect of either the strong or the comprehensive mandates on completed fertility.

In sum, our paper shows that, despite being associated with higher birth rates among relatively older women and with a higher prevalence of multiple births, infertility insurance does not have a statistically significant effect on women's fertility at the end

<sup>7</sup> Although researchers paid attention to differential impacts on current and completed fertility in the 1980s (e.g., [Ward and Butz 1980](#)), we are not aware of any other paper addressing the effects of infertility insurance on women's completed fertility.

<sup>8</sup> For the first exercise, we construct the probability of having a biological child by the age of 30 using data from the June Marriage and Fertility Supplement of the Current Population Survey (the 'June CPS'). For the second exercise, we combine birth certificate data from the National Vital Statistics with data from the March Annual Social and Economic Supplement of the Current Population Survey (the 'March CPS') and apply the synthetic control group method ([Abadie et al. 2010](#)), which relies on more general identifying assumptions than the standard difference-in-differences model typically used in the literature.

<sup>9</sup> These results refer to a set of six states that enacted what is usually labeled as "strong-to-cover" infertility mandates, defined in Sect. 2. They are: Arkansas, Hawaii, Illinois, Maryland, Massachusetts, and Rhode Island.

<sup>10</sup> In an independent work concurrent with our own, [Ohinata \(2011\)](#) offers an alternative to [Buckles \(2005\)](#) based on the estimation of a duration model for age at first birth using longitudinal data from the panel study of income dynamics (PSID). She finds a substantial delay of motherhood of approximately 1.5–2 years. Ohinata's identification is, however, based on a relatively small number of women.

<sup>11</sup> The prevalence of multiple births is approximately 31 % in ART cycles using fresh non-donor eggs or embryos ([CDC 2007](#)) compared with slightly more than 3 % in the rest United States population.

**Table 1** Infertility treatment mandates classifications

State	Cover/offer	Mandatory IVF coverage	Application	Marriage required
<b>Arkansas</b>	<b>Cover-strong (1987)</b>	<b>Yes</b>	<b>HMOs excluded</b>	<b>Yes</b>
California	Offer (1989)	No	All plans	No
Connecticut	Offer (1989)	Yes	HMOs excluded	No
<b>Hawaii</b>	<b>Cover-strong (1987)</b>	<b>Yes</b>	<b>All plans</b>	<b>Yes</b>
<b>Illinois</b>	<b>Cover-strong (1991)</b>	<b>Yes</b>	<b>All plans</b>	No
<b>Maryland</b>	<b>Cover-strong (1985)</b>	<b>Yes</b>	<b>All plans</b>	<b>Yes</b>
<b>Massachusetts</b>	<b>Cover-strong (1987)</b>	<b>Yes</b>	<b>All plans</b>	No
Montana	Cover-weak (1987)	No	HMOs only	No
New York	Cover-weak (1990)	No	HMOs excluded	No
Ohio	Cover-weak (1991)	No	HMOs only	No
<b>Rhode-Island</b>	<b>Cover-strong (1989)</b>	<b>Yes</b>	<b>All plans</b>	No
Texas	Offer (1987)	Yes	All plans	Yes
West Virginia	Cover-weak (1977)	No	HMOs only	No

Cover strong are indicated in bold

Sources: [Buckles \(2005\)](#), [Schmidt \(2007\)](#) and the National Infertility Association (<http://www.resolve.org/>). Louisiana and New Jersey enacted infertility mandates in 2001, but these states were excluded from our analysis

of their reproductive lives. The reason lies, as we further show, in the fact that infertility insurance mandates also appear to delay motherhood among relatively younger women and, hence, make conception more difficult because fecundity decreases with age.

The rest of the paper is structured as follows: Sect. 2 describes the characteristics of infertility treatment mandates including where and when they were enacted; Sect. 3 describes the data sources used in this paper; Sect. 4 presents our evidence on the delay of motherhood; Sect. 5 presents an analysis of the impact of the mandates on women's completed fertility; Sect. 6 presents conclusions; Sect. 7 contains figures and tables; and Sect. 8 is the "Appendix".

## 2 Infertility treatment mandates

Table 1 summarises the major features of infertility insurance mandates and their timing. The classification of mandates is consistent with those presented in [Buckles \(2005\)](#) and [Schmidt \(2007\)](#). Mandates can either require mandatory coverage of infertility treatment for all plans ('mandates to cover') or demand that employers offer at least one plan that covers infertility treatment ('mandates to offer'). In addition, mandates to cover are described as 'strong' when they cover IVF treatment and at least 35 % of the women are affected by the mandate; otherwise they are described as 'weak'.<sup>12</sup> According to the American Society for Reproductive Medicine, of the

<sup>12</sup> The classification of mandates as being 'strong' versus 'weak' is not universal, although it is broadly consistent with the classifications used in the literature and is adopted here for convenience. In contrast to [Schmidt \(2007\)](#), [Buckles \(2005\)](#) cites Ohio as having a non-IVF coverage mandate.

six states classified as ‘mandate-to-cover-strong’, only Arkansas does not apply the mandate to all plans (HMOs are exempt). In addition, out of the six strong mandate states, three require women to be married to benefit from the insurance coverage (see Mookim et al. 2008 for more detail on mandates).

Other authors, such as Hamilton and McManus (2012), Bundorf et al. (2007), and Mookim et al. (2008), classify Illinois, Massachusetts, and Rhode-Island (hereafter IL-MA-RI) as having ‘universal’, ‘comprehensive’, and the ‘most comprehensive coverage’, respectively. In this paper, the effects of infertility treatment mandates on this specific group of states are also analysed.

Concerns about policy endogeneity have been discussed in previous studies, such as Bitler and Schmidt (2012) and Hamilton and McManus (2012). The latter two studies in particular reached the conclusion that the enactment of infertility treatment mandates was largely the result of the efforts of a national infertility association (RESOLVE) and the political rather than fertility preferences of state residents. We briefly revisit this issue in our sample in Sect. 4.1.

The “Appendix” describes state-specific changes made to the original strong mandates in later periods. Because most of the revisions that occurred within the sample period (i.e., before 2001) undercut benefits, they are expected to decrease the estimated effects of the mandates.

### 3 Data sources

Our data comes from three main sources: (1) birth certificates from the National Vital Statistics System of the National Center for Health Statistics; (2) the March Annual Social and Economic Supplement of the Current Population Survey (March CPS) and (3) the June Marriage and Fertility Supplement of the Current Population Survey (June CPS).<sup>13</sup>

In Sect. 4.1, we use individual-level data from the June CPS to analyse the impact of the mandates on the probability of having a child by the age of 30/35. The period of analysis is restricted to 1979–1995 due to the unavailability of the main dependent variable beyond that period.

In Sect. 4.2, we aggregate individual-level data from birth certificates on the age of new mothers at the state and year level to analyse the impact of the mandates on the age at first birth. The birth certificates contain individual records for 50 % of the births occurring within the United States from 1968 to 1971; from 1972 to 1984, the data are based on 100 % samples of birth certificates from some states and on 50 % samples from the remaining states; as of 1985, the data cover every birth from all reporting areas.<sup>14</sup> These data also contain information on the mother, including her age, race and state of residence as well as specific information about the timing, parity (whether the birth was a first or subsequent birth), and plurality (the number of children per delivery

<sup>13</sup> We downloaded March CPS data and documentation from the IPUMS-USA database (King et al. 2010), while we used processed June CPS files from Unicon Research Corporation (<http://www.unicon.com>).

<sup>14</sup> Births occurring to US citizens outside the United States are not included. The number of states from which 100 % of the records are used increased from 6 in 1972 to all states and the District of Columbia in 1985. We adjusted the total numbers accordingly in the analysis.

that is, whether it was a single, twin, triplet, or higher-order birth) of each birth. This information allows the identification of first births and, therefore, the determination of the average age of new mothers, which is the main variable of interest in Sect. 4.2. When multiple births occur, only one observation per delivery is kept in our sample to avoid oversampling multiple-birth mothers, who are more likely to be older and/or to have used ARTs.<sup>15</sup>

The birth certificate data also contain other potentially relevant socioeconomic variables, such as marital status and maternal education, but the information is not always complete or available throughout the sample period.<sup>16</sup> Thus, for the multivariate analyses in Sect. 4.2, we combine the aggregate state and year level birth certificate information on the age of new mothers with a richer set of socioeconomic characteristics obtained from the March CPS, including race, education, marital and labour market status, wages, and health insurance coverage. Note that controlling for employment-sponsored health insurance coverage is important in this context given that uninsured individuals are not directly affected by the mandates and that most non-elderly insured individuals in the United States obtain insurance through their workplace.<sup>17</sup>

Our analysis in Sect. 4.2 could, in principle, be conducted until 2005, when state identifiers become unavailable in the natality files. However, we restrict our sample to the period before 2001 (i.e., from 1972 to 2001) so that Louisiana and New Jersey can be included as controls (these two states passed infertility insurance laws in 2001). Including Louisiana and New Jersey in the treated group would not have provided us with sufficient post-intervention years to analyse the long-term impact of these latest mandates. Moreover, because states were not uniquely identified in the CPS until 1977, the controls from the March CPS are only included from 1977 onwards. The period of analysis, unless otherwise stated, is 1972–2001, inclusively. To further enrich the set of control variables, state-year legal abortion rates by 1000 women aged 15–44 and state of residence obtained from The Guttmacher Institute are included.

<sup>15</sup> We uniquely identify multiple-birth mothers by using, whenever available, various variables, such as the year, month and day of birth, the gestation time, the state, county and place or facility of birth, the presence of an attendant at birth, plurality; maternal age, race, years of schooling, marital status, the place of birth, the state, county, city, and standard metropolitan statistical area (SMSA) of residence; and paternal age and race.

<sup>16</sup> Importantly, information on maternal education is missing for the following states and years: California (1972–1988), Alabama (1972–1975), Arkansas (1972–1977), Connecticut (1972), the District of Columbia (1972), Georgia (1972), Idaho (1972–1977), Maryland (1972–1973), New Mexico (1972–1979), Pennsylvania (1972–1975), Texas (1972–1988) and Washington (1972–1991). Marital status is not reported in any state until 1978.

<sup>17</sup> An important feature of state-mandated benefits is that self-insured employers are exempt from state insurance regulations under the 1974 Federal Employee Retirement Income Security Act (ERISA). Hence, employers who self-insure are exempt from the requirements of the state infertility insurance mandates previously described. Because self-insured companies are typically large, the impact of the mandates is likely to be concentrated on small firms. Lacking information on the self-insured status of employers, researchers have used firm size as a proxy for ERISA exemptions (e.g., Schmidt 2007 and the references therein, Simon 2004; Bhattacharya and Vogt 2000). Self-reported firm-size from the March CPS could be used as a proxy for ERISA exemption status, but this variable was not recorded before 1988 and therefore could not be included as a predictor in our estimations.



Finally, in Sect. 5, we again use data from the June CPS to estimate the effects of the mandates on completed fertility using information for 1979–2000 as well as for the extended period 1979–2008 as a robustness check. Unlike the March CPS data, which is available on a yearly basis and only provides information on the presence of children in the household without discriminating between biological and non-biological children, the June questionnaire is not administered every year but contains information on the number of biological children ever born. In particular, the June CPS provides this information for the following years during our sample period: 1979–1985, 1990–1992, 1994–1995, 1998, and 2000.<sup>18</sup> Additionally, the June CPS contains information on other potential determinants of fertility, such as age, marital status, and labour market status, which we incorporate as controls in the regressions. An important handicap of both the June and March CPS questionnaires for our purposes is that only information about the current state of residence is given and not, for example, information on the birth state of a child.

#### 4 The effect of infertility treatment mandates on the delay of motherhood

In this section, we provide evidence that infertility insurance mandates cause a strategic delay of motherhood. We start in Sect. 4.1 by showing that relatively young women are less likely to have children after the mandates. Although this exercise is similar to Buckles' (2005), we use more adequate data and a different empirical specification. The premise in Buckles (2005) is that insurance for infertility treatment allows women to postpone motherhood and invest in their careers.<sup>19</sup> Buckles uses the enactments of mandates to cover infertility treatment in several states of the United States during the late 1980s and 1990s as natural experiments and reports that the mandates increased the probability that relatively old women (40–49) would have young children. We refer to this non-strategic response as a short-term compositional effect or as ex-post moral hazard.<sup>20</sup> Additionally, Buckles (2005) shows that the mandates reduced the probability that relatively young women (22–26 and 26–30) would have young children. Showing a decrease in the probability of having children while young, however, falls short of proving delay because it fails to consider what these women do when they grow older. Indeed, these women could decide to remain childless. To demonstrate delay, one needs to combine this evidence with evidence of the timing of first-time motherhood, i.e., when (if at all) women decide to become mothers and stop delaying motherhood. This is exactly what we do in Sect. 4.2. The idea is as follows: if young women postpone childbearing because of the mandates, then we should see

<sup>18</sup> In 1977, 1986, 1987, and 1988, the 'number of babies' question was also asked but only to women who had ever been married. The question is most often posed to women in their childbearing years, which in the June CPS typically included women ages 18–44. Including women aged 45–49 would limit the analysis to the years 1979, 1983, 1985, and 1995, leaving us with very few post-intervention periods.

<sup>19</sup> In a very recent paper Kroeger and La Mattina (2015) revisited the link between infertility treatment and career development.

<sup>20</sup> Other papers providing evidence in favour of a non-strategic effect are Schmidt (2005a, 2007) and Bitler and Schmidt (2012).

an increase in the average age at first birth several years after the policies are implemented. We compute the average age of first-time mothers at the state-year level from birth certificate data and show that mandates not only increase the average age in the short-run—consistent with a compositional effect—but, more importantly, that the first motherhood age-gap relative to the counterfactual increases with time since enactment—consistent with a behavioural effect. The counterfactual in this exercise is constructed using the synthetic control method recently developed by [Abadie et al. \(2010\)](#) (henceforth ADH), which exhibits several advantages over the conventional DID estimator. The ADH methodology is explained in Sect. 4.2 in detail.

#### 4.1 The probability of having a first child by 30 and 35 years of age

In this section, we estimate the effect of time since the enactment of a mandate on the probability of having at least one biological child by the ages of 30 and 35 using the June CPS data.

Table 2 presents the estimated marginal effects from probit estimations of the number of years of mandated coverage at age 30 on the probability of having at least one biological child by that age. In all regressions, women are, by definition, older than 30. The first four columns show the marginal effects for all strongly treated states, whereas the last four columns show the marginal effects for IL-MA-RI. Each set of four columns is further split between the effects for ‘all’ women and the effects for ‘whites’ only. Finally, the table presents the results of regressions where the control group is composed of all non-treated states (columns labelled ‘control’) and of regressions where the control group is restricted to the states belonging to the synthetic control group constructed in Sect. 4.2 with the ADH methodology (columns labelled ‘synth’; see the note to Table 2 for a complete list of states in this control group).<sup>21</sup> Panel A shows the marginal effects when the independent variables of interest are year intervals since enactment by the age of 30, i.e., ‘1–5’ and ‘6–10 years’. Panel B shows the marginal effects when the independent variable of interest is instead expressed as a quadratic polynomial of years since enactment by the age of 30. In a given state and year, the number of years of mandated coverage by 30 varies by women according to age. For example, a woman from Maryland who turned 30 before 1985 had zero years of mandated coverage by 30; a woman from Maryland who turned 30 in 1990 had 5 years of mandated coverage at age 30 but would have had 10 years of mandated coverage by 30 had she turned 30 in 1995. Therefore, the coefficient on the variable ‘1–5 years of mandated coverage at age 30’ is being identified by relatively older women, while the coefficient on the variable ‘6–10 years of mandated coverage at age 30’ is being identified by the younger cohorts. Additionally, note that because states enacted their mandates in different years, the number of years of mandated coverage by a certain age is not collinear with age (e.g., a woman who experienced 5 years of mandated coverage by age 30 in Illinois would be 6 years younger than a woman from Maryland with the same duration of coverage by age 30). A large set of controls

<sup>21</sup> We thank an anonymous referee for proposing to use the states belonging to the synthetic control group of Sect. 4.2 as the control group for the results in this section as well.

**Table 2** Marginal effects of infertility insurance mandates on the probability of having at least one child by age 30

	Strong states				IL-MA-RI			
	All		Whites		All		Whites	
	Control (1)	Synth (2)	Control (3)	Synth (4)	Control (5)	Synth (6)	Control (7)	Synth (8)
Panel A: years of mandated coverage								
1–5 years	0.0040 (0.0071)	0.0062 (0.0100)	0.0144* (0.0081)	0.0221** (0.0108)	0.0115*** (0.0044)	0.0162** (0.0081)	0.0210*** (0.0050)	0.0303*** (0.0076)
6–10 years	−0.0290*** (0.0098)	−0.0285** (0.0124)	−0.0188 (0.0116)	−0.0115 (0.0140)	−0.0335*** (0.0074)	−0.0297** (0.0137)	−0.0223*** (0.0082)	−0.0135 (0.0149)
Pseudo R <sup>2</sup>	0.118	0.126	0.128	0.137	0.118	0.139	0.128	0.153
Log-lik.	−50,006	−24,497	−42,699	−20,381	−47,396	−17,777	−40,997	−14,428
Panel B: quadratic polynomial specification								
Evaluated at 5 years	−0.0054 (0.0038)	−0.0056 (0.0039)	−0.0070* (0.0040)	−0.0069* (0.0041)	−0.0124*** (0.0011)	−0.0129*** (0.0013)	−0.0133*** (0.0018)	−0.0135*** (0.0021)
Evaluated at 10 years	−0.0134 (0.0133)	−0.0145 (0.0140)	−0.0237* (0.0127)	−0.0266** (0.0132)	−0.0381*** (0.0036)	−0.0419*** (0.0042)	−0.0452*** (0.0063)	−0.0497*** (0.0065)
Pseudo R <sup>2</sup>	0.118	0.126	0.128	0.137	0.118	0.139	0.128	0.154

**Table 2** continued

	Strong states		Whites		IL-MA-RI		Whites	
	All		Control		All		Control	
	Control (1)	Synth (2)	Control (3)	Synth (4)	Control (5)	Synth (6)	Control (7)	Synth (8)
Log-lik.	−50,006	−24,497	−42,699	−20,382	−47,395	−17,777	−40,996	−14,428
No. of obs	109,211	51,557	93,121	42,346	103,499	37,785	89,337	30,474
<i>P</i> treated	0.685		0.675		0.661		0.652	

Results from Probit estimation. The dependent variable is 1 if a woman has at least one child by age 30 and zero otherwise. In Panel A, the independent variables of interest are dummies of year-intervals (e.g. 1–5 and 6–10 years) since the mandates were enacted by the age of 30. In Panel B, the independent variable of interest is a quadratic polynomial of the number of years since enactment by the age of 30. Sample years: 1979–1983, 1985, 1990, 1992, 1995. The age range is 30–44 both inclusive. Control group in columns labeled “control” is composed of all non-treated control states. Control group in columns labeled “synth” i.e. columns (2), (4), (6) and (8) is restricted to states used to construct the Synthetic Control Group in Sect. 4.2. of the paper. Hence, in col. (2) control states are Alaska, Arizona, District of Columbia, Michigan, Minnesota, Nevada, New Jersey, North Carolina, North Dakota, South Carolina, Utah, Vermont and Wisconsin. In col. (4), control states are Arizona, Colorado, District of Columbia, Michigan, Minnesota, Nevada, New Jersey, North Carolina, North Dakota, South Carolina, Utah, Vermont and Wisconsin. In col. (6), control states are Alabama, Louisiana, Michigan, Minnesota, New Jersey, Utah, Vermont and Wyoming. In col. (8), control states are Colorado, District of Columbia, Louisiana, Michigan, Minnesota, New Jersey, Utah, Vermont and Wyoming. All regressions control for year dummies, state fixed effects, education variables (high-school, more than high-school), working status, unmarried status and age dummies. Regressions for all women also include race dummies. “P treated” stands for the average probability of at least one child before 30 among the treated. Levels of statistical significance: \*\*\* denotes significance at the 1-% level; \*\* at the 5-% level; and \* at the 10-% level. Standard errors are robust and clustered at the state level

are included in all regressions, such as state fixed effects, year fixed effects, educational attainment dummy variables (viz., high school, beyond high school), working status, married status, and age dummy variables in 5-year intervals. Standard errors are clustered at the state level.

The results in Panel A of Table 2 show that having a strong or comprehensive mandate enacted by age 30 for 1–5 years is associated with a higher probability of having at least one child before age 30, and this effect is statistically significant for the strong mandates ('whites' sample) and for the comprehensive mandates (both for 'all' and for 'whites'). As noted above, the 1–5 years effect is identified by the relatively older cohorts, who did not act strategically and who increased their fertility due to the mandates. This result, which is consistent with that of [Bundorf et al. \(2007\)](#), is suggestive of a moral hazard effect among relatively fertile couples.<sup>22</sup> However, facing a strong or comprehensive mandate by the age of 30 for longer than 6 years is associated with a lower probability of having a biological child by 30. The marginal effects for '6–10 years' are identified by the relatively younger cohorts, and, hence, constitute evidence of strategic delay. The magnitudes of these marginal effects are, as expected, larger for IL-MA-RI and are generally statistically significant. Moreover, the marginal effects are not small in magnitude, as they imply a reduction of 2.9–3.3 percentage points (pp) in the probability of having a child by age 30 for all women (representing a decrease of 4.2–5 % in the average probability of the treated presented at the bottom of Table 2) and between 1.1 and 2.2 pp for white women (or a decrease of 1.7–3.4 % in the average probability of the treated presented at the bottom of Table 2).

The fit obtained with the quadratic polynomial specification (Panel B) is essentially identical to the one obtained in Panel A, as can be observed from the values of the log-likelihood. The marginal effects of the mandates are now consistently negative when evaluated at 5 and 10 years after their enactment and are always statistically significant, except for the strong-mandates 'all'-women sample. The quadratic specification also delivers more intuitive results; for example, the effect for 'whites' is always larger than the effect for 'all'. The magnitudes of the marginal effects evaluated at 10 years are also quite large, especially for the strong-mandates 'whites' sample as well as for the comprehensive mandates.

A general concern with policy evaluation studies based on non-experimental designs is the possibility that such studies are flawed because the adoption of policies is often endogenous. This would be the case if, for example, the enactment of infertility treatment mandates was linked to low fertility or to a systematic pattern of motherhood delay. As briefly explained in Sect. 2, the endogeneity of the infertility treatment mandates is rejected by several authors ([Bitler and Schmidt 2012](#); [Hamilton and McManus 2012](#); [Abramowitz 2014](#)). Nonetheless, because our approach and variables are different from theirs, we assess the extent to which endogeneity might be an issue in our own data. Similar to [Bitler and Schmidt \(2012\)](#), we have included leads of the mandate variables in our analysis of the probability of at least one child by age 30. In particular, we have experimented with a linear measure of 'years to enactment of mandate' as

<sup>22</sup> An anonymous referee suggested an alternative explanation where some low fertility couples can conceive very quickly after IVF.

well as with indicator variables for ‘future mandate’ together with indicator dummies for 1–3 or 1–5 years to the enactment of the mandates (these results are not displayed for the sake of brevity). None of these variables is statistically significantly different from zero.

Finally, Table 3 reports the marginal effects of the probability of having at least one child by age 35. The results displayed show that most marginal effects are not statistically significantly different from zero, indicating that delay is no longer statistically significant by the age of 35.

Our estimations, although similar, offer some advantages over those of Buckles (2005). First, our dependent variable does not suffer from two important shortcomings present in her indicator ‘own children younger than six present in the household’. Her variable, constructed from the March CPS, does not distinguish between biological and non-biological children, and its universe is restricted to children who live in the household. To construct our variable, we use data from the June CPS on biological children ever born. Second, in our case, the relevant number of years since the mandates is measured at age 30, and all women in the sample are, by definition, older than 30. Buckles (2005), uses the number of years since the mandates at the time of the interview in a sample which includes women who were children—as young as 8-year-old—at the time the mandates were enacted. Her linear specification forces the coefficients on the interaction terms between the years since the mandates and age to be close to zero and not significant because a 22-year-old woman who experiences a mandate for 14 years, for example, is hardly more affected than a 22-year-old who has only been under a mandate for 2 years. Unfortunately, our approach comes at a cost: the variable from the June CPS used to construct the dependent variable used in the regressions of Tables 2 and 3 was not recorded beyond 1995, thereby restricting the estimation of the effects of the mandates to the medium run.<sup>23</sup>

Using the March CPS proxy ‘age of own eldest child in the household’, which was recorded until 2008, to increase the number of available periods would be unlikely to bias the results for relatively young women (whose children have not yet left the household and whose probability of adoption is smaller), but it would likely bias the results for older women whose eldest child has already left the household. Indeed, the latter could be wrongly classified as having zero children by the ages of 30/35. Moreover, older women are also more likely to have zero years of mandated coverage by 30/35 (the omitted category in the regressions of Panel A). Consequently, using the March CPS data would imply mistakenly assigning less relative fertility by age 30/35 to older women who have had less exposure to the mandates by the age of 30 or 35, causing an upward bias in the estimated effect of the number of years of coverage on the probability of having at least one child by ages 30/35. Buckles (2005) estimates of

<sup>23</sup> To construct the variable ‘at least one child by age  $x$ ’, where we use  $x = 30$  and  $x = 35$  years of age, we need to know the age at which each woman had her first child. We obtain this information from the June CPS variable ‘birth1y’, which reports the year of birth of the first child. Unfortunately, this variable is not recorded every year and is not available beyond 1995. Moreover, for the years 1990, 1992, and 1995, ‘birth1y’ contains many missing values, although they are evenly spread across all states. The number of missing values for ‘birth1y’ is approximately 53 and 55 % for the control and the strongly treated states, respectively.

**Table 3** Marginal effects of infertility insurance mandates on the probability of having at least one child by age 35

	Strong states				IL-MA-RI			
	All		Whites		All		Whites	
	Control (1)	Synth (2)	Control (3)	Synth (4)	Control (5)	Synth (6)	Control (7)	Synth (8)
Panel A: years of mandated coverage								
1–5 years	−0.0230*	−0.0221*	−0.0198	−0.0202	−0.0199	−0.0174	−0.0164	−0.0135
	(0.0120)	(0.0132)	(0.0137)	(0.0140)	(0.0176)	(0.0187)	(0.0171)	(0.0176)
6–10 years	−0.0127	−0.0094	−0.0033	−0.0055	0.0043	0.0073	0.0160	0.0161
	(0.0100)	(0.0117)	(0.0149)	(0.0163)	(0.0151)	(0.0168)	(0.0102)	(0.0111)
Pseudo R <sup>2</sup>	0.117	0.126	0.131	0.143	0.119	0.137	0.131	0.162
Log-lik.	−24,769	−12,335	−21,091	−10,144	−23,465	−8,859	−20,247	−7,028
Panel B: quadratic polynomial specification								
Evaluated at 5 years	0.0010	0.0014	0.0028	0.0023	0.0017	0.0014	0.0041	0.0030
	(0.0022)	(0.0022)	(0.0036)	(0.0034)	(0.0032)	(0.0032)	(0.0027)	(0.0026)
Evaluated at 10 years	0.0120	0.0121	0.0158	0.0148	0.0105	0.0081	0.0154	0.0113
	(0.0084)	(0.0085)	(0.0119)	(0.0108)	(0.0136)	(0.0134)	(0.0123)	(0.0116)
Pseudo R <sup>2</sup>	0.117	0.126	0.131	0.143	0.118	0.137	0.131	0.162
Log-lik.	−24,770	−12,336	−21,091	−10,144	−23,466	−8,860	−20,248	−7,029
No. of obs	67,618	32,008	57,953	26,506	63,971	23,273	55,498	18,805
P treated	0.766		0.761		0.743		0.738	

Results from Probit estimation. The dependent variable is 1 if a woman has at least one child by age 30 and zero otherwise. In Panel A, the independent variables of interest are dummies of year-intervals (e.g. 1–5 and 6–10 years) since the mandates were enacted by the age of 35. In Panel B, the independent variable of interest is a quadratic polynomial of the number of years since enactment by the age of 35. Sample years: 1979–1983, 1985, 1990, 1992, 1995. The age range is 35–44 both inclusive. Control group in columns labeled “control” is composed of all non-treated control states. Control group in columns labeled “synth” i.e. columns (2), (4), (6) and (8) is restricted to states used to construct the Synthetic Control Group in Sect. 4.2. of the paper. Hence, in col. (2) control states are Alaska, Arizona, District of Columbia, Michigan, Minnesota, Nevada, New Jersey, North Carolina, North Dakota, South Carolina, Utah, Vermont and Wisconsin. In col. (4), control states are Arizona, Colorado, District of Columbia, Michigan, Minnesota, Nevada, New Jersey, North Carolina, Virginia, Washington, Wisconsin and Wyoming. In col. (6), control states are Alabama, Alaska, District of Columbia, Louisiana, Michigan, Minnesota, New Jersey, Utah, Vermont and Wyoming. In col. (8), control states are Colorado, District of Columbia, Louisiana, Michigan, Minnesota, New Jersey, Utah, Vermont and Wyoming. All regressions control for year dummies, state fixed effects, education variables (high-school, more than high-school), working status, unmarried status and age dummies. Regressions for all women also include race dummies. “P treated” stands for the average probability of at least one child before 35 among the treated. Levels of statistical significance: \*\*\* denotes significance at the 1-% level; \*\* at the 5-% level; and \* at the 10-% level. Standard errors are robust and clustered at the state level

the probability of the ‘presence of small children in the household’ for older women (Table 4 of her paper) are likely to suffer from this bias.

It is important to recognise that the June and March CPS data share a shortcoming due to the lack of information on past states of residence. This limitation would be problematic if women with infertility issues were more likely to travel to mandated states to pay lower prices for infertility treatments. [Abramowitz \(2014\)](#) claims that this is unlikely because interstate migration during 1981–2010 was only 3 %, and mandated states in general had lower immigration than non-mandated states. Finally, some readers may wonder whether the welfare reform enacted in 1996 [the Personal Responsibility and Work Opportunity Act (PRWORA)] affects or biases our results; therefore in the “Appendix”, we explain why it does not.

## 4.2 Average maternal age at first birth

We begin this part of the analysis by providing some descriptive evidence that broadly characterises the patterns of fertility timing across groups of states and over time. Figure 2a–c plot the evolution of the average age of new mothers in control states versus all treated states, all strongly treated states, and IL-MA-RI (the states with ‘comprehensive coverage’), respectively. The two vertical lines in each figure indicate the years in which the first and last of the corresponding mandates were passed for all of the treated states (1977; 1991), for the strongly treated states (1985; 1991) and for IL-MA-RI (1987; 1991). Although the average age of first-time mothers was higher in treated states than in control states even before any mandate was enacted, Fig. 2b, c show that, for states with ‘strong mandates to cover’ and for IL-MA-RI, the treated-control gap became larger after the passage of the mandates. For example, in 2001, the age gap between IL-MA-RI and the control states was slightly more than 16 months, that is, nearly 4 months longer than in 1991, the year in which the latest strong mandate passed in Illinois.<sup>24</sup>

More interestingly, from the viewpoint of this paper, the observed increase in the treated-control gap is statistically significant at standard levels of testing, which suggests that the effect of the mandates is larger in the long run than in the short run.<sup>25</sup> The observed treated-control gap follows an analogous pattern, and its magnitude is similar, albeit somewhat larger, when the sample is restricted to white mothers.<sup>26</sup> It is also worth noting that the increasing trend that we have documented is not so

<sup>24</sup> These 4 months account for 33 % of the overall increase in the age of new mothers that occurred in IL-MA-RI between 1991 and 2001. Although this relative magnitude is purely descriptive, the fact that it is so large further motivates our subsequent analysis.

<sup>25</sup> In particular, we regressed maternal age at first birth for the period 1972–2001 from birth certificates from the National Vital Statistics System dataset on a set of state and year fixed effects and on indicators of the number of years passed since the mandates were enacted in the mother’s state of residence. Subsequently, we tested whether the effect of the number of years since the passage of mandates was statistically significantly larger in the long run than in the short run. Our results indicate that this was indeed the case for the ‘strong mandates to cover’ and the ‘comprehensive mandates’. See [Machado and Sanz-de-Galdeano \(2011\)](#) for a more detailed discussion of these results.

<sup>26</sup> These results are available in an earlier working paper version of this article ([Machado and Sanz-de-Galdeano 2011](#)).



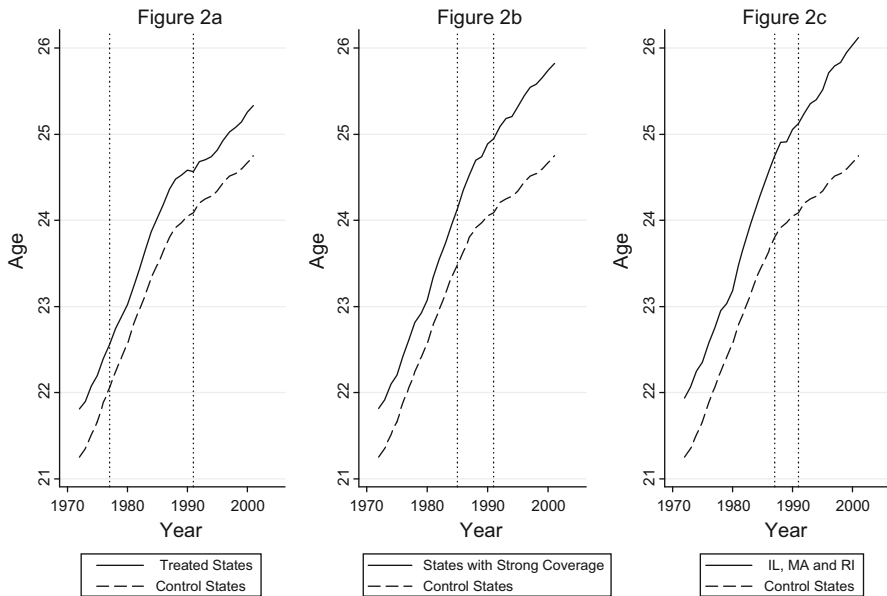
**Table 4** Means of predictors used in synthetic control group estimation for maternal age at first birth

	States with Strong coverage	Control group	
		Synthetic	Control states
% married women (1982–1984)	0.52436	0.52577	0.55883***
Abortion rate (1978–1982)	29.1949	31.8491***	24.4055***
% white females (1982–1984)	0.81651	0.81951	0.84786***
% white females (1977–1981)	0.82268	0.82580	0.85711***
% black females (1981–1984)	0.13906	0.14113	0.13267**
% black females (1977–1980)	0.13639	0.13620	0.12836***
% highly-educated women (1982–1984)	0.36872	0.36441	0.31839***
% highly-educated women (1977–1981)	0.31365	0.31367	0.28049***
Female employment rate (1982–1984)	0.61664	0.61537	0.59480***
Female participation rate (1977–1984)	0.64740	0.64819	0.63591***
Previous year female log hourly wage (1982–1984)	1.95119	1.94975	1.84945***
Previous year female employment rate (1983–1984)	0.65491	0.65574	0.64217**
Previous year female employment rate (1977–1982)	0.63055	0.62794	0.62046***
% of women covered by ESI in own name (1982–1984)	0.34135	0.36550***	0.33498
Maternal age at first birth, 1984	23.9426	23.9471	23.3413***
Maternal age at first birth, 1982	23.5474	23.5310	22.9629***
Maternal age at first birth, 1981	23.3371	23.3320	22.7918***
Maternal age at first birth, 1979	22.9212	22.9489	22.4131***
Maternal age at first birth, 1977	22.6122	22.6055	22.0579***
Maternal age at first birth, 1976	22.4229	22.4288	21.8879***
Maternal age at first birth, 1975	22.2059	22.2042	21.6589***
Maternal age at first birth, 1974	22.0984	22.0898	21.5136***
Maternal age at first birth, 1973	21.9157	21.9237	21.3490***
Maternal age at first birth, 1972	21.8149	21.8108	21.2541***
% of new mothers age >35 (1981–1983)	0.02131	0.02164	0.01516***
% of new mothers age >35 (1977–1980)	0.01396	0.01436**	0.00978***

Treatment group composed of new mothers in strong coverage states

Columns (1) and (2) are obtained directly from the synthetic group estimation routine. Column (3) was added for comparison. The control group in column (3) is composed of all control states. The predictors obtained from the Natality data are: “Maternal age at first birth (YEAR)” and the “Mean percent of new mothers age >35 (YEAR INTERVAL)”. All the other predictors are obtained from the March CPS data. The sample used for the predictors from the March CPS are women between 15- and 49-year-old. Each predictor variable is averaged for the period(s) indicated. ESI stands for employment sponsored health insurance. Composition of the synthetic control group varies with the sample and can be checked in the notes of the following tables

evident when all of the thirteen treated states are considered together (Fig. 2a). The reason lies in the much more limited scope of the ‘weak mandates to cover’ and the ‘mandates to offer’, described in Sect. 2. Interestingly, visual inspection of Fig. 2 indicates that the treated-control gap may have been increasing even before the mandates were enacted, especially for IL-MA-RI. Although this evidence is merely



**Fig. 2** Average maternal age at first birth. All women

descriptive, it highlights the importance of selecting a control group that successfully mimics the dynamics of the treated states to estimate the true impact of the mandates.

To construct a control group that maximises the similarities between women in treated and control states, we use the synthetic control method (Abadie et al. 2010)<sup>27</sup> which benefits from several advantages over the conventional DID estimator. The synthetic control group approach limits the discretion of researchers in the choice of the control units by offering a procedure for the construction of an ‘ideal’ control group denoted as the ‘synthetic’ control group. The synthetic control group uses a weighted average of the potential control units, which provides a better counterpart for the treated units than any single actual control unit or set of actual control units. The weights assigned to each control unit are chosen to minimise the differences in pre-treatment trends and other predictors between the treated unit and the synthetic control group. This estimation procedure is very transparent because it reports the estimated relative contribution, which may be zero, of each control unit to the synthetic group. It is worth noting that, although the synthetic control group approach is obviously related to the standard DID estimator, which it extends, the synthetic control group approach also has features in common with matching estimators insofar as both approaches attempt to minimise observable differences between the treatment and control units. Indeed, some of the latest developments in the literature attempt to minimise the chances of

<sup>27</sup> See Abadie and Gardeazabal (2003) for an earlier application of the synthetic control group approach.

selection into treatment based on unobservables.<sup>28</sup> The synthetic control approach is a step in this direction because it relies on more general identifying assumptions than the standard DID model, allowing the effects of unobserved variables on the outcome to vary with time.

To apply the synthetic control group, the birth certificate data on the age of new mothers for the period 1972–2001 must be aggregated at the state and year levels.<sup>29</sup> This aggregation is advantageous in our case because it allows us to control for socioeconomic characteristics by merging the aggregated birth certificate data with socioeconomic variables available in the March CPS for the period 1977–2001 (also aggregated at the state and year levels).<sup>30</sup> Moreover, births from all strongly treated states are also aggregated as if they belonged to the same state with initial treatment in the year 1985, the year the first strong mandate was enacted. Similarly, we aggregated the data for the subset of comprehensive states (IL-MA-RI) with initial treatment in the year 1987 when the first mandate was enacted in Massachusetts.<sup>31</sup> The synthetic control group is constructed as the convex combination of control states that are most similar to the states with strong coverage and comprehensive coverage with respect to various socioeconomic predictors as well as lagged values of the average age of first motherhood before treatment (i.e., before 1985 and 1987, respectively). More precisely, the predictors chosen include the following: (1) variables that control for the demographic and family structure of the female population, such as the percentage of new mothers older than 35 and the percentage of married women in the state; (2) variables that control for the state's race composition, such as percentage of white and black females; (3) variables that control for the education level of the female population, such as the percentage of highly educated women; (4) variables related to the female labour market, such as the participation rate and employment rate, the average logarithm of the hourly wage, and the percentage of women covered by Employment Sponsored Insurance (ESI); (5) variables that control for differences in abortion laws or attitudes, such as the abortion rate per 1000 women by state of residency; and (6)

<sup>28</sup> These concerns have been raised in several studies (e.g., Heckman et al. 1997, 1998; Michalopoulos et al. 2004; Smith and Todd 2005), where it was argued that matching on observables alone would not guarantee an adequate counterfactual because unobservables may affect the selection into treatment thereby leading to bias in the estimation of treatment effects. Heckman et al. (1997, 1998) and Smith and Todd (2005) present evidence that highlights the advantages of using a DID matching strategy, which allows for time-invariant differences between the treatment and control groups. Michalopoulos et al. (2004) allows for selection into treatment based on individual-specific unobserved linear trends.

<sup>29</sup> Note that the synthetic control group methodology may not be used with individual-level data.

<sup>30</sup> Note that the analysis allows one to rely on data covering different periods. Hence, we use data on the average age of new mothers for the period 1972–2001 and exogenous characteristics, from the March CPS, for the period 1977–2001. This implies that, before 1977, only the average age of first-time mothers in different years was used as predictors, while after 1977 a richer set of predictors has been included.

<sup>31</sup> To ensure our results are not driven by the assumption of the initiation of treatment in 1985 and 1987 for strongly treated states and IL-MA-RI, respectively, we performed a simple but extreme robustness test that consists of attributing the treatment year to the year the last state enacted the mandate. This implies that the treatment year becomes 1991 for both the strongly treated and the comprehensive states. Because there are states in both groups that passed their mandates before 1991, this would result in understating the effect of the mandates (i.e., the estimated effect may be regarded as a lower bound). Our estimated effect is, as expected, somewhat lower for all samples considered—ranging from 59 to 97 %—but remains significantly positive.

several lags of average age at first birth.<sup>32</sup> All of these predictors are averaged over different periods to maximise the fit of the estimation. Although the predictors are roughly the same for the four estimations (strong, strong whites only, IL-MA-RI, IL-MA-RI whites only), the composition of the synthetic control group is not exactly the same. It is always the case, however, that New Jersey is systematically the most important state in the composition of the four synthetic control groups, representing between 26 and 41 %, followed by Minnesota, whose contribution ranges between 11 and 17 % of the estimated synthetic control group.<sup>33</sup>

Table 4 presents the pre-treatment (i.e., before 1985) sample averages of all predictors for the states with strong coverage (column 2), as well as for the synthetic control group (column 3), and for the full group of control states (column 4). As shown, prior to the passage of the first strong mandate to cover, new mothers in control states were already younger than in states where strong mandates to cover eventually passed. These mothers also earned lower wages on average and, were less educated, more likely to be married, less likely to have an abortion, less likely to participate in the labour market, less likely to be employed and less likely to have employer-provided health insurance coverage. The predictors' pre-treatment values for the strongly treated states resemble the pre-treatment values of the synthetic control group (column 3) much more than the pre-treatment values for the full set of control states (column 4). Hence, the synthetic control group should be a better counterfactual for the treated groups. Tables for the white sample and for IL-MA-RI are similar but are not reported here in the interest of brevity.

Our synthetic control estimate of the impact of the infertility coverage mandates on the timing of the first child is the difference between the average age of new mothers in states with strong mandates to cover (or the subset of IL-MA-RI) and the synthetic control group at a given date. Panel A of Table 5 shows the estimates for the group of states with strong mandates to cover, whereas Panel B shows the same estimates for IL-MA-RI. The second column reports the synthetic control group estimate in 2001, that is, 16 and 10 years after the first and the last strong mandates were passed, respectively. We refer to this estimate as the long-term effect of the mandates. For the group of states with strong mandates, the long-term effect amounts to 0.266 and 0.317 years, approximately 3.2 months for all women and 3.8 months for white women, respectively. For IL-MA-RI, the effects are larger despite the shorter period since the first mandate: an increase of approximately 4.1–5.4 months in the average age at first child for all and for white new mothers, respectively. The estimated long-term effects of the mandates are considerable—between 15.7 and 18.8 % of the total increase from 1985 to 2001 for the group with strong coverage and between 24.8 and

<sup>32</sup> Other variables were considered as predictors but were discarded because they worsen the fit of the model, i.e., they increased the root mean squared prediction error (rmspe) of the estimation, which is a measure of the difference between the treated and the synthetic control group during the pre-treatment period. These variables include, for example, the average number of children in the household, the split of the female population's age structure into 5-year age brackets, the percentage of females with private health insurance, the percentage of first-deliveries in different 5-year age brackets, the average company size for female workers, and the year of divorce reforms according to Friedberg (1998) and Gruber (2004).

<sup>33</sup> Results for the composition of the four synthetic control groups are not reported for the sake of brevity but are available from the authors upon request.

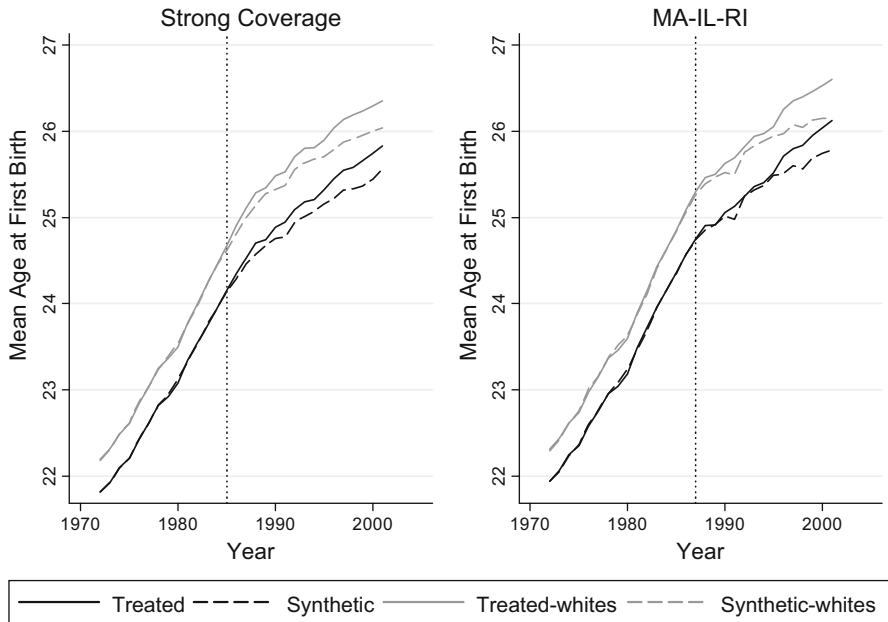
**Table 5** The long-run impact of strong infertility insurance coverage mandates on the age of new mothers

	Synthetic control group estimates		
	Parameter estimate (2001)	rmspe	p value of rmspe ratio
Panel A: strong mandates			
(1) All	0.266	0.0177	0.026**
(2) Whites	0.317	0.0179	0.053*
Panel B: Illinois, Massachusetts and Rhode Island			
(3) All	0.341	0.0237	0.079*
(4) Whites	0.448	0.0269	0.053*

Treatment is assumed to start in 1985 for states with strong mandates and in 1987 for IL, MA and RI. rmspe denotes the root mean squared prediction error. All the p values displayed are based on placebo runs that are described in Sect. 4.2. The states that enacted strong mandates are: Arkansas (1987), Hawaii (1987), Illinois (1991), Maryland (1985), Massachusetts (1987) and Rhode-Island (1989). Predictors used in estimation of the synthetic control effect are described in Table 4 for the all women sample in strong mandated states. For other samples, the tables were omitted for the sake of brevity but the set of predictors is the same. Detailed tables are available in Machado and Sanz-de-Galdeano (2011). The composition of the synthetic control group varies with the sample. Hence, in row (1) control states are Alaska, Arizona, District of Columbia, Michigan, Minnesota, Nevada, New Jersey, North Carolina, North Dakota, South Carolina, Utah, Vermont and Wisconsin. In row (2), control states are Arizona, Colorado, District of Columbia, Michigan, Minnesota, Nevada, New Jersey, North Carolina, Virginia, Washington, Wisconsin and Wyoming. In row (3), control states are Alabama, Alaska, District of Columbia, Louisiana, Michigan, Minnesota, New Jersey, Utah, Vermont and Wyoming. In row (4), control states are Colorado, District of Columbia, Louisiana, Michigan, Minnesota, New Jersey, Utah, Vermont and Wyoming. Levels of statistical significance: \*\*\* denotes significance at the 1-% level; \*\* at the 5-% level; and \* at the 10-% level. The estimates in this section were obtained using the October 2011 version of SYNTH, the Stata module to implement synthetic control methods programmed by Abadie, Diamond and Hainmueller (see <http://ideas.repec.org/c/boc/bocode/s457334.html>)

34.3 % for IL-MA-RI. The synthetic control estimates are slightly smaller than the raw DID aggregate estimate, which amounts to approximately 0.42 years (5 months). The third column of Table 5 shows the root mean squared prediction error (rmspe), which is a measure of the difference in age at first birth between the treated and the synthetic control group during the pre-treatment period. Hence, the lower the rmspe, the better is our counterfactual. The rmspe values, displayed in column 3, are all small, demonstrating the good fit of the models.

Inference in the synthetic control estimation method is often non-standard because the number of non-treated units is typically small. However, as ADH argue in both their 2010 and their 2014 papers, “by systematizing the process of estimating the counterfactual of interest, the synthetic control method enables researchers to conduct a wide array of falsification exercises” or “placebo studies” that can be used for inference. We follow this approach and apply the synthetic control method to every potential control state to create distributions of 38 placebo treatment effects and other statistics. ADH recommend using the resulting distribution of the ratio post/pre-intervention rmspe values to construct a p value for this statistic. The p value is constructed by simply calculating the proportion of the estimated placebo ratios of post/pre-intervention rmspe values that are greater than or equal to the ratio for the truly treated states. The

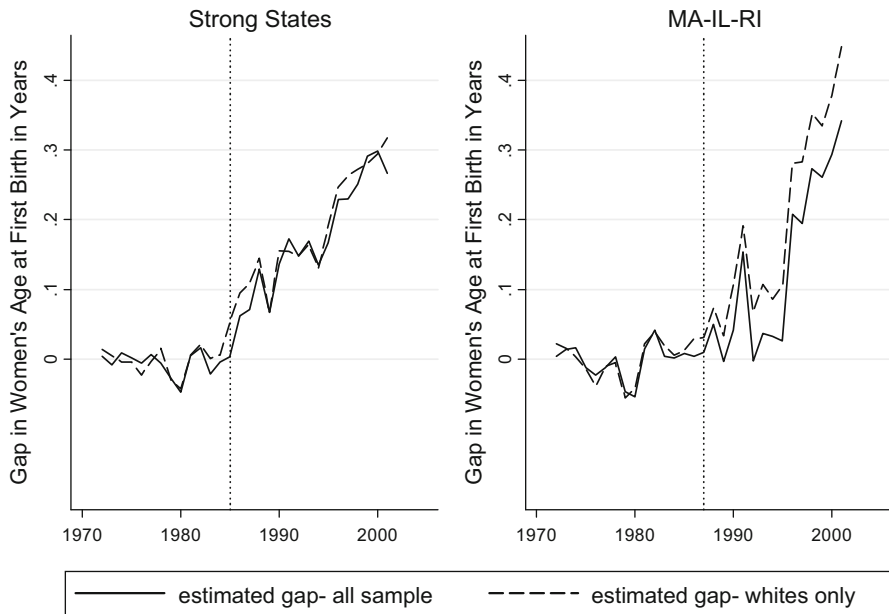


**Fig. 3** Average maternal age at first birth: treated vs. synthetic control groups

idea is that, in the absence of a treatment effect, the ratio of the post/pre-intervention fit should be similar for treated and non-treated units. As the last column in Table 5 shows, the  $p$  values for the post/pre-intervention rmspe are all very small, implying the existence of a statistically significant treatment effect for all four samples.

Figure 3 shows the annual average age at first birth in strongly treated states and in IL-MA-RI compared with the synthetic control group counterpart for the sample period (1972–2001) for all women and for white women. The synthetic control group does a good job in tracking the pre-treatment evolution of new mothers' ages in states with strong coverage and in IL-MA-RI, which indicates we have a good approximation to the counterfactual trend in maternal age at first birth that states with strong or comprehensive coverage would have experienced had the mandates not been enacted. It is worth noting the contrast with the evolution of all non-treated units used in Fig. 2b, c, which has failed to track the treated states' pattern as closely as the synthetic control group has done. This result is not surprising, given the low rmspe values and the closeness in terms of predictor values between the states with strong coverage and their synthetic version shown in Table 4, for example.

More important than the size of the estimated long-term effect is its evolution over time, which is shown in Fig. 4. Regressions of the estimated annual effects of the 17 post-treatment periods for the strong mandate states (and the 15 post-treatment periods for IL-MA-RI) on indicators of time since the mandates (i.e., less than 5 years since the mandate, between 6 and 10 years, or more than 10 years), shown in Table 6, confirm that the impact of the mandates grew significantly over time. Figure 4 and Table 6 are crucial because they demonstrate that the long-term cumulative



**Fig. 4** Gap between treated and synthetic groups for age at first birth

impact of the mandates on the timing of first births extends beyond its short-term non-strategic impact on older women with infertility problems. The increasing impact of the mandates together with the results from Sect. 4.1 constitute evidence of the delay of motherhood. We believe that the mechanism operating here is simple; suppose no supply constraints existed for infertility treatments when the mandates were enacted. If mandates had only a non-strategic effect on older women (i.e., ex-post moral hazard), the estimated effect should therefore be positive but nearly constant over time. The long-term effect may be larger than the short-term effect because, for example, women who were young when the mandates were enacted strategically delay motherhood (i.e., exert ex-ante moral hazard). An alternative explanation for the increasing effect may be that supply constraints for fertility treatments existed when the mandates were enacted but gradually disappeared due to technological improvements and/or price reductions, giving access to a larger number of users of infertility treatments. Our analysis cannot identify the exact contribution of each of these potential explanations to the increasing effect of the mandates.<sup>34</sup>

<sup>34</sup> We interpret an increase in the number of people who have insurance coverage as a decrease in prices and, hence, as a supply shock. We thank an anonymous referee for this example. Another explanation for the growing gap, noted by an anonymous referee, would be the following: suppose only highly educated women delay motherhood irrespective of ART coverage. If the trend in the share of highly educated women differed in the treated states relative to the synthetic control group, then a growing share of older women would look for treatment in the treated states and the compositional effect would be long-lived. However, we take into account the pre-treatment trend in the percentage of highly educated women in the construction of the synthetic control group. Therefore, only major changes in the female educational trends after treatment could cause a long-lived compositional effect.

**Table 6** Evolution of the synthetic control gap in maternal age at first birth. OLS estimates

	Strong mandates		IL, MA, RI	
	All	Whites	All	Whites
Mandated coverage				
1–5 years	0.093*** (0.016)	0.114*** (0.016)	0.048 (0.029)	0.095*** (0.027)
6–10 years	0.158*** (0.007)	0.158*** (0.010)	0.100** (0.041)	0.172*** (0.045)
More than 10 years	0.261*** (0.012)	0.279*** (0.010)	0.292*** (0.017)	0.378*** (0.024)
F-test of joint significance	[<0.001]	[<0.001]	[<0.001]	[<0.001]
t-tests of equality				
“1–5” vs. “6–10” coeff.	[0.002]	[0.036]	[0.324]	[0.163]
“6–10” vs. “more than 10” coeff.	[<0.001]	[<0.001]	[0.0011]	[0.0017]
“1–5” vs. “more than 10” coeff.	[<0.001]	[<0.001]	[<0.001]	[<0.001]
No. obs.	17	17	15	15
R <sup>2</sup>	0.982	0.980	0.881	0.924

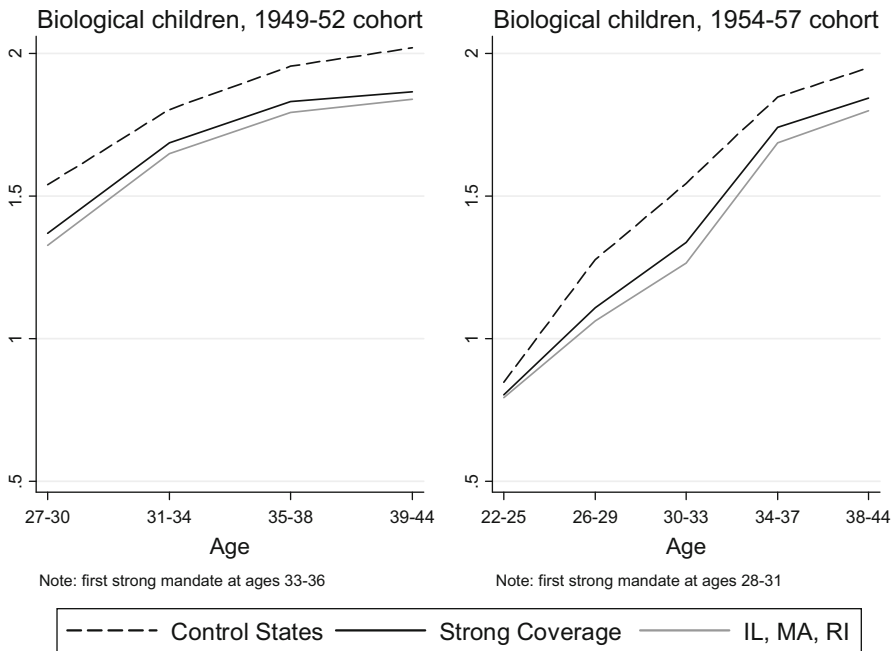
The dependent variable is the post-treatment estimated gap of maternal age at first birth between states with strong coverage and the synthetic control group. These variables correspond to those plotted in Fig. 4. There are 17 post-treatment periods, corresponding to 1985–2001 for the strongly treated states and 15 post-treatment periods for IL-MA-RI, corresponding to the period 1987–2001. The synthetic control group has different compositions depending on the sample used. Hence, for strong mandates “All”, the synthetic control group is composed of: Alaska, Arizona, District of Columbia, Michigan, Minnesota, Nevada, New Jersey, North Carolina, North Dakota, South Carolina, Utah, Vermont and Wisconsin. For strong mandates “whites”, the synthetic control group is composed of: Arizona, Colorado, District of Columbia, Michigan, Minnesota, Nevada, New Jersey, North Carolina, Virginia, Washington, Wisconsin and Wyoming. For IL-MA-RI “All”, the synthetic control group is composed of: Alabama, Alaska, District of Columbia, Louisiana, Michigan, Minnesota, New Jersey, Utah, Vermont and Wyoming. For IL-MA-RI, “White”, the synthetic control group is composed of: Colorado, District of Columbia, Louisiana, Michigan, Minnesota, New Jersey, Utah, Vermont and Wyoming. Levels of statistical significance: \*\*\* denotes significance at the 1-% level; \*\* at the 5-% level; and \* at the 10-% level. p values corresponding to the F-tests of joint significance and the one-sided t-tests of equality are displayed in square brackets. Robust standard errors are displayed in round brackets. The model includes no constant

## 5 The effect of infertility treatment mandates on completed fertility

In the previous section, we presented evidence that infertility treatment mandates induced ex-ante moral hazard leading women to delay motherhood. However, would delaying motherhood necessarily result in a lower number of children per woman? There are at least two elements that may operate in opposite directions. On the one hand, by inducing delay, mandates may negatively affect the total number of pregnancies per woman; on the other hand, the higher probability of multiple births amongst patients of infertility treatments may compensate for any negative effect on the number of deliveries.<sup>35</sup> In this section, we estimate the effect of infertil-

<sup>35</sup> For example Gumus and Lee (2012) find that adoption decreases the number of ART cycles.





**Fig. 5** Average number of own biological children for two cohorts in June CPS

ity insurance coverage mandates on the total number of biological children per woman.

Figure 5 plots the average number of biological children over a woman's reproductive life for two cohorts (women born between 1949–1952 and between 1954–1957) in the control, strongly treated and comprehensive states (IL-MA-RI) using data from the June CPS.<sup>36</sup> When the first strong mandate was enacted in Maryland in 1985, women in the older (younger) cohort were between 33–36 (28–31) years old. The figure shows that women in strongly treated and comprehensively treated states have, on average, a smaller number of biological children and do not catch up with women in the control group by the age of 44. In our estimations below, we account for differences in observable characteristics between treated and control states in two ways. First, we control for covariates that may affect the trends shown in Fig. 5. Second, we restrict the control group in the estimation to states that have a positive weight in the construction of the synthetic control group of Sect. 4.2.<sup>37</sup>

We estimate the effects of time passed since the mandates were enacted on the total number of biological children using a zero-inflated Poisson regression for 44-year-old women, i.e., women at the end of their reproductive lives, controlling for

<sup>36</sup> In the June CPS, the number of biological children was obtained systematically only from women who were 44-year-old or younger. Although some women have children beyond the age of 44, these women are very few in number.

<sup>37</sup> We thank an anonymous referee for this suggestion.

**Table 7** The effect of infertility insurance mandates on the number of biological children

	Strong coverage against				MA-IL-RI against			
	Control states		Synth states		Control states		Synth states	
	All	Whites	All	Whites	All	Whites	All	Whites
Mandated coverage	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1–5 years	0.133 (0.145)	0.120 (0.119)	0.117 (0.167)	0.139 (0.120)	0.114 (0.181)	0.129 (0.150)	0.151 (0.230)	0.203 (0.265)
6–10 years	–0.076 (0.122)	–0.078 (0.136)	–0.045 (0.139)	–0.052 (0.137)	–0.068 (0.165)	–0.124 (0.164)	–0.044 (0.179)	–0.068 (0.172)
More than 10 years	–0.047 (0.147)	–0.039 (0.146)	0.006 (0.149)	–0.009 (0.159)	–0.075 (0.100)	–0.086 (0.088)	–0.096 (0.120)	–0.144 (0.138)
% of zeros	13.11	13.10	13.72	14.08	13.21	13.14	13.75	13.66
Vuong test p value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Log-likelihood	–15,322	–12,871	–7026	–5645	–14,557	–12,356	–5259	–4129
No. obs.	8609	7365	3966	3260	8163	7068	2938	2357

Marginal effects from zero-inflated Poisson. Sample years until 2000

Data from the June Marriage and Fertility Supplement of the Current Population Survey. Women aged 44 only. Sample years 1979–1985, 1990, 1992, 1994, 1995, 1998, 2000. Control group in columns (3), (4) and (7), (8) is restricted to states used to construct the Synthetic Control Group in Sect. 4.2. of the paper. Hence, in col. (3) control states are Alaska, Arizona, District of Columbia, Michigan, Minnesota, Nevada, New Jersey, North Carolina, North Dakota, South Carolina, Utah, Vermont and Wisconsin. In col. (4), control states are Arizona, Colorado, District of Columbia, Michigan, Minnesota, Nevada, New Jersey, North Carolina, Virginia, Washington, Wisconsin and Wyoming. In col. (7), control states are Alabama, Alaska, District of Columbia, Louisiana, Michigan, Minnesota, New Jersey, Utah, Vermont and Wyoming. In col. (8), control states are Colorado, District of Columbia, Louisiana, Michigan, Minnesota, New Jersey, Utah, Vermont and Wyoming. The control group in other columns includes all non-treated states i.e. states where mandates were not enacted during the sample period. All models include year and state fixed effects as well as educational attainment indicators, a binary variable indicating whether the woman works or not and a unmarried dummy variable. Regressions for all women include race dummies as well. Levels of statistical significance: \*\*\* denotes significance at the 1-% level, \*\* at the 5-% level, and \* at the 10-% level. Standard errors, displayed in round brackets, are clustered at the state level

individual characteristics.<sup>38</sup> Table 7 shows the estimated marginal effects of time since the mandates for sample years 1979–2000 (including Louisiana and New Jersey in the control group as in the previous sections). The first four columns of Table 7 show the marginal effects for all strongly treated states, whereas the last four columns show the marginal effects for IL-MA-RI. Each set of four columns is further split into two groups of results. The first group is obtained from regressions in which the control group is composed of all non-treated states (the first two columns); the second group

<sup>38</sup> To further assess which model best suits our data, we compare the Akaike’s information criterion (AIC) values of several models, namely Poisson, negative binomial, generalized Poisson, generalized negative binomial, zero-inflated Poisson and zero-inflated negative binomial. The model with the lowest AIC value both for the whole sample and for the sample of whites is the zero-inflated Poisson. The zero-inflated Poisson model allows for a different process for the zeros and accommodates excess frequency of zeros relative to the Poisson model. The results with all models described above were qualitatively similar.

**Table 8** The effect of infertility insurance mandates on the number of biological children

	Strong coverage against				MA-IL-RI against			
	Control states		Synth states		Control states		Synth states	
	All	Whites	All	Whites	All	Whites	All	Whites
Mandated coverage	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1–5 years	0.1392 (0.1298)	0.1285 (0.1175)	0.1210 (0.1461)	0.1372 (0.1250)	0.1259 (0.1631)	0.1190 (0.1405)	0.1866 (0.2013)	0.2185 (0.2135)
6–10 years	−0.0808 (0.1166)	−0.0871 (0.1318)	−0.0948 (0.1411)	−0.0595 (0.1440)	−0.0670 (0.1558)	−0.1272 (0.1550)	−0.0530 (0.1927)	−0.0761 (0.1915)
More than 10 years	−0.0401 (0.0906)	−0.0290 (0.0814)	0.0142 (0.1079)	−0.0066 (0.0851)	−0.0438 (0.1054)	−0.0847 (0.0891)	−0.0587 (0.1163)	−0.0318 (0.1088)
% of zeros	15.08	15.09	15.97	16.05	15.06	15.07	15.95	16.05
Vuong test p value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Log-likelihood	−19,642	−16,470	−8597	−6917	−18,574	−15,802	−5743	−4584
No. obs.	11,335	9691	4978	4100	10,698	9274	3286	2673

Marginal effects from zero-inflated Poisson. Sample years until 2008

Data from the June Marriage and Fertility Supplement of the Current Population Survey. Women aged 44 only. Sample years 1979–1985, 1990, 1992, 1994, 1995, 1998, 2000, 2002, 2004, 2006, 2008. Because Louisiana and New Jersey were treated in 2001, i.e. during the sample years, we drop these two states from the sample. The control group in columns (3), (4) and (7), (8) is restricted to states used to construct the Synthetic Control Group in Sect. 4.2. of the paper, except for Louisiana and New Jersey. Hence, in col. (3) control states are Alaska, Arizona, District of Columbia, Michigan, Minnesota, Nevada, North Carolina, North Dakota, South Carolina, Utah, Vermont and Wisconsin. In col. (4), control states are Arizona, Colorado, District of Columbia, Michigan, Minnesota, Nevada, North Carolina, Virginia, Washington, Wisconsin and Wyoming. In col. (7), control states are Alabama, Alaska, District of Columbia, Michigan, Minnesota, Utah, Vermont and Wyoming. In col. (8), control states are Colorado, District of Columbia, Michigan, Minnesota, Utah, Vermont and Wyoming. The control group in other columns includes all non-treated states i.e. states where mandates were not enacted during the sample period. All models include year and state fixed effects as well as educational attainment indicators, a binary variable indicating whether the woman works or not and a unmarried dummy variable. Regressions for all women include race dummies as well. Levels of statistical significance: \*\*\* denotes significance at the 1-% level, \*\* at the 5-% level, and \* at the 10-% level. Standard errors, displayed in round brackets, are clustered at the state level

is obtained from regressions in which the control group is restricted to states used in the construction of the synthetic control group of Sect. 4.2 (the last two columns). The table shows the results for all women and white women separately. Standard errors are clustered at the state level. Finally, the p-values of the Vuong test show that the zero-inflated Poisson model is preferred (in all cases) to the Poisson model.

The relatively large positive impact in the first five years under the mandates, although not statistically significant, plunges to negative values in the following years. None of these effects statistically differs significantly from zero, however, due to their large estimated standard errors. To improve efficiency, we replicate the estimation by extending the sample up to 2008 (see Table 8).<sup>39</sup> Consistent with the rest of the

<sup>39</sup> Importantly, note that the dependent variable used in this section for complete fertility—i.e., ‘biological children ever born’—is available from the June CPS until 2008, contrary to the case of ‘age at first child’ used in Sect. 4.1, which is only available until 1995.

**Table 9** The effect of infertility insurance mandates on the number of biological children

	Strong coverage against				MA-IL-RI against			
	Control states		Synth states		Control states		Synth states	
	All	Whites	All	Whites	All	Whites <sup>(a)</sup>	All	Whites
Quadratic evaluated at	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
5 years	−0.0058 (0.0125)	−0.0052 (0.0146)	−0.0051 (0.0138)	−0.0045 (0.0150)	−0.0170* (0.0103)	−0.0155 (0.0146)	−0.0202** (0.0102)	−0.0205** (0.0093)
10 years	−0.0162 (0.0205)	−0.0107 (0.0241)	−0.0178 (0.0231)	−0.0148 (0.0265)	−0.0460 (0.0463)	−0.0201 (0.0410)	−0.0659 (0.0540)	−0.0817 (0.0593)
15 years	−0.0266 (0.0465)	−0.0162 (0.0527)	−0.0304 (0.0519)	−0.0251 (0.0556)	−0.0751 (0.1000)	−0.0247 (0.0906)	−0.1117 (0.1138)	−0.1430 (0.1221)
Log-likelihood	−15,325	−12,843	−7028	−5637	−14,558	−12,607	−5261	−4131
No. obs.	8609	7365	3966	3260	8163	7069	2938	2357

Quadratic polynomial specification. Marginal effects from zero-inflated Poisson. Sample years until 2000. Data from the June Marriage and Fertility Supplement of the Current Population Survey. Women aged 44 only. Sample years 1979–1985, 1990, 1992, 1994, 1995, 1998, 2000. The control group in columns (3), (4) and (7), (8) is restricted to states used to construct the Synthetic Control Group of Sect. 4.2 of the paper. Hence, in col. (3) control states are Alaska, Arizona, District of Columbia, Michigan, Minnesota, Nevada, New Jersey, North Carolina, North Dakota, South Carolina, Utah, Vermont and Wisconsin. In col. (4), control states are Arizona, Colorado, District of Columbia, Michigan, Minnesota, Nevada, New Jersey, North Carolina, Virginia, Washington, Wisconsin and Wyoming. In col. (7), control states are Alabama, Alaska, District of Columbia, Louisiana, Michigan, Minnesota, New Jersey, Utah, Vermont and Wyoming. In col. (8), control states are Colorado, District of Columbia, Louisiana, Michigan, Minnesota, New Jersey, Utah, Vermont and Wyoming. The control group in other columns includes all non-treated states i.e. states where mandates were not enacted during the sample period. All models include year and state fixed effects as well as educational attainment indicators, a binary variable indicating whether the woman works or not and a unmarried dummy variable. Regressions for all women include race dummies as well. (a) means that one state dummy was dropped from the inflated part of the model in order to achieve convergence. Levels of statistical significance: \*\*\* denotes significance at the 1-% level, \*\* at the 5-% level, and \* at the 10-% level. Standard errors, displayed in round brackets, are clustered at the state level

paper, when presenting results using data until 2008, we exclude Louisiana and New Jersey from the sample because these states enacted mandates in 2001. Results with the sample years until 2008 remain qualitatively the same, and none of the marginal effects of time since enactment are statistically significant.<sup>40,41</sup>

Conceivably, the large standard errors obtained in Tables 7 and 8 may result from poor specification. We re-estimated the model using polynomials of the years since enactment instead of the time intervals ‘1–5’, ‘6–10’, and ‘more than 10’ used in Tables 7 and 8. Tables 9 and 10 show the results for the quadratic specification. The log-likelihood with the quadratic specification is slightly worse than that obtained with

<sup>40</sup> If we include New Jersey among the strongly treated states when using the data until 2008, we obtain slightly larger effects, where the short-run effect of 1–5 years is marginally statistically significant at the 10 % level for the sample of whites only.

<sup>41</sup> A linear model of the total number of biological children on year intervals since the mandates—i.e., ‘1–5’, ‘6–10’ and ‘more than 10’—also produced non-statistically significant effects.

**Table 10** The effect of infertility insurance mandate on the number of biological children

	Strong coverage against				MA-IL-RI against			
	Control states		Synth states		Control states		Synth states	
	All	Whites	All	Whites	All	Whites	All	Whites
Quadratic evaluated at	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
5 years	−0.0041 (0.0121)	−0.0060 (0.0135)	−0.0050 (0.0129)	−0.0065 (0.0137)	−0.0030 (0.0152)	−0.0073 (0.0159)	−0.0062 (0.0150)	−0.0046 (0.0165)
10 years	−0.0047 (0.0035)	−0.0014 (0.0049)	−0.0018 (0.0045)	−0.0017 (0.0057)	−0.0118*** (0.0043)	−0.0110** (0.0049)	−0.0163** (0.0075)	−0.0131 (0.0092)
15 years	−0.0053 (0.0141)	0.0031 (0.0167)	0.0014 (0.0128)	0.0031 (0.0181)	−0.0205 (0.0215)	−0.0147 (0.0218)	−0.0264 (0.0252)	−0.0215 (0.0295)
Log-likelihood	−19,644	−16,471	−8599	−6918	−18,575	−15,803	−5742	−4585
No. obs.	11,335	9691	4978	4100	10,698	9274	3286	2673

Quadratic polynomial specification. Marginal effects from zero-inflated Poisson. Sample years until 2008. Data from the June Marriage and Fertility Supplement of the Current Population Survey. Women aged 44 only. Sample years 1979–1985, 1990, 1992, 1994, 1995, 1998, 2000, 2002, 2004, 2006, 2008. Because Louisiana and New Jersey were treated in 2001, i.e. during the sample years, we have dropped these two states from the sample. The control group in columns (3), (4) and (7), (8) is restricted to states used to construct the Synthetic Control Group in Sect. 4.2. of the paper, except for Louisiana and New Jersey. Hence, in col. (3) control states are Alaska, Arizona, District of Columbia, Michigan, Minnesota, Nevada, North Carolina, North Dakota, South Carolina, Utah, Vermont and Wisconsin. In col. (4), control states are Arizona, Colorado, District of Columbia, Michigan, Minnesota, Nevada, North Carolina, Virginia, Washington, Wisconsin and Wyoming. In col. (7), control states are Alabama, Alaska, District of Columbia, Michigan, Minnesota, Utah, Vermont and Wyoming. In col. (8), control states are Colorado, District of Columbia, Michigan, Minnesota, Utah, Vermont and Wyoming. The control group in other columns includes all non-treated states i.e. states where mandates were not enacted during the sample period. All models include year and state fixed effects as well as educational attainment indicators, a binary variable indicating whether the woman works or not and a unmarried dummy variable. Regressions for all women include race dummies as well. Levels of statistical significance: \*\*\* denotes significance at the 1-% level, \*\* at the 5-% level, and \* at the 10-% level. Standard errors, displayed in round brackets, are clustered at the state level

the time intervals, but the values are generally similar. The estimated marginal effects at 5, 10 and 15 years after the mandates are small and statistically insignificant for strongly treated states, with some statistically negative values for IL-MA-RI. Because the quadratic specification has little flexibility to accommodate the ups and downs of the marginal effects within the first few years, we re-estimate the model using a cubic specification, not shown for the sake of brevity. The cubic specification, however, produces marginal effects of the same order of magnitude, where even more terms are statistically negative for the case of IL-MA-RI. Thus far, our results do not show evidence of a positive effect of infertility treatment mandates on total fertility.

Finally, we estimated the effects of years since the mandates on total fertility conditional on being a mother, i.e., we drop the zeros from the estimation. Table 11 shows the marginal effects from a truncated Poisson regression for the sample up to 2000. Not surprisingly, we find that the marginal effects on the intensive margin are smaller in magnitude and also not statistically significantly different from zero.

Overall, the results show no statistically significant effect of either the strong or the comprehensive mandates on completed fertility.

**Table 11** The effect of infertility insurance coverage mandates on the number of biological children conditional on being a mother

	Strong coverage against				MA-IL-RI against			
	Control states		Synth states		Control states		Synth states	
	All	Whites	All	Whites	All	Whites	All	Whites
Mandated coverage								
1–5 years	0.0650 (0.1526)	0.0805 (0.1679)	0.0577 (0.1649)	0.0565 (0.1815)	0.0870 (0.1673)	0.0861 (0.1772)	0.0950 (0.1867)	0.0684 (0.2049)
6–10 years	−0.0217 (0.1671)	−0.0386 (0.1844)	−0.0381 (0.1783)	−0.0552 (0.1958)	0.0290 (0.2016)	−0.0639 (0.2129)	0.0150 (0.2196)	−0.0995 (0.2380)
More than 10 years	0.0751 (0.2394)	−0.0271 (0.2718)	0.0785 (0.2537)	−0.0344 (0.2855)	−0.0594 (0.3802)	−0.0432 (0.3832)	−0.1173 (0.4002)	−0.1205 (0.4137)
Log-likelihood	−12,286.5	−10,276.8	−5598.8	−4468.2	−11,660.3	−9888.4	−4197.7	−3315.9
No. obs.	7480	6400	3422	2801	7085	6139	2534	2035

Marginal effects from truncated Poisson model. Sample years until 2000

Data from the June Marriage and Fertility Supplement of the Current Population Survey. Women aged 44 only. Sample years 1979–1985, 1990, 1992, 1994, 1995, 1998, 2000. The control group in columns (3), (4) and (7), (8) is restricted to states used to construct the Synthetic Control Group in Sect. 4.2. of the paper. Hence, in col. (3) control states are Alaska, Arizona, District of Columbia, Michigan, Minnesota, Nevada, New Jersey, North Carolina, North Dakota, South Carolina, Utah, Vermont and Wisconsin. In col. (4), control states are Arizona, Colorado, District of Columbia, Michigan, Minnesota, Nevada, New Jersey, North Carolina, Virginia, Washington, Wisconsin and Wyoming. In col. (7), control states are Alabama, Alaska, District of Columbia, Louisiana, Michigan, Minnesota, New Jersey, Utah, Vermont and Wyoming. In col. (8), control states are Colorado, District of Columbia, Louisiana, Michigan, Minnesota, New Jersey, Utah, Vermont and Wyoming. The control group in other columns includes all non-treated states i.e. states where mandates were not enacted during the sample period. All models include year and state fixed effects as well as educational attainment indicators, a binary variable indicating whether the woman works or not and a unmarried dummy variable. Regressions for all women include race dummies as well. Levels of statistical significance: \*\*\* denotes significance at the 1-% level, \*\* at the 5-% level, and \* at the 10-% level. Standard errors, displayed in round brackets, are clustered at the state level

## 6 Conclusions

In this paper we ask two questions about the impact of infertility treatment insurance coverage on women’s fertility. First, does an increase in the coverage of infertility treatments lead to greater delay of motherhood? Second, does the coverage of infertility treatment increase completed fertility by the end of a woman’s reproductive life? Variation in the timing of the enactment of infertility insurance mandates during the late 1980s and early 1990s across the United States is exploited to answer the two questions.

We focus our analysis on the effect of ‘strong mandates to cover’, which were passed in six states in the United States (the treatment group): Arkansas (1987), Hawaii (1987), Illinois (1991), Maryland (1985), Massachusetts (1987), and Rhode-Island (1989), as well as on a subset of this group that enacted the most comprehensive mandates (Illinois, Massachusetts, and Rhode-Island). Our analysis uses birth certificate data from the National Vital Statistics, data from the March Annual Social and Economic Sup-

plement of the Current Population Survey (March CPS), and data from the Marriage and Fertility Supplement of the Current Population Survey (June CPS). Our results indicate that, despite the fact that infertility insurance coverage has been shown to increase birth rates among relatively older women as well as the prevalence of multiple births, strong or comprehensive mandates have no statistically significant and positive effect on completed fertility. This is because infertility insurance mandates also appear to delay motherhood among relatively younger women and hence make conception more difficult due to the fact that fecundity decreases with age. In particular, we find evidence in favour of a behavioural response in the form of a delay of motherhood in the states that have enacted strong and comprehensive mandates both through a reduction of the probability of having a child by the age of 30 and through an increasing effect of the mandates on the average age at first birth over time. Therefore, our results suggest that the coverage of infertility treatments, which has been considered a potential policy tool to increase European fertility rates, would not contribute to a long-term increase in fertility.

**Acknowledgments** We are grateful to Alberto Abadie, Manuel Arellano, Manuel Bagües, Marianne Bitler, Giorgio Brunello, Guillermo Caruana, Monica Costa-Dias, Joao Ejarque, Roger Feldman, Eugenio Giolito, Libertad González, Rachel Griffith, Andrea Ichino, Julián Messina, Pedro Mira, Enrico Moretti, Nuno Sousa Pereira, Helena Szrek, Marcos Vera, Ernesto Villanueva and two anonymous referee for helpful comments as well as the audiences at the 4th COSME Workshop on Gender Economics (2011), ASHEcon (2010), Health Econometrics Workshop (2010), ESPE (2009), Barcelona GSE Trobada (2009), 3rd PEJ conference, EARIE (2009), SAE (2009), and APES (2009). Matilde P. Machado acknowledges financial support from the Spanish Ministry of Science and Technology Grant SEJ2007-66268 and Ministerio de Economía y Competitividad Grant ECO2010-20504. Anna Sanz-de-Galdeano is also affiliated with CRES-UPF, IZA and MOVE. She acknowledges financial support from the Spanish Ministry of Science and Technology Grants ECO2011-28822 and ECO2014-58434-P.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendix 1: Could the welfare reform explain our results?

Other potential causes may exist for the growing gap between treated and synthetic states. For example, during the post-treatment years either treated or control states may have enacted laws—e.g. welfare reform—that affected the mean age at first birth. Welfare reform is likely to have discouraged maternity at younger ages through its demanding work requirements and stringent eligibility standards for acceptance into the assistance programs and, therefore, could have increased the mean age at first birth.<sup>42</sup> The welfare reform was enacted in 1996 [Personal Responsibility and Work Opportunity Act (PRWORA)] and became effective in July 1997, four years before the end of our sample. Before that, however, some states had already introduced work

<sup>42</sup> To our knowledge, no study has found that welfare reform increased the age of first motherhood. In fact, Hao and Cherlin (2004) compare two cohorts of young women and conclude that welfare reform has not decreased teenage fertility.

requirements for welfare eligibility in the 1980s and early 1990s (Mead 2004). For the interpretation of our analysis it is important to know the group (treatment, control, or not in the sample) to which these early adopters of welfare reform belong. If some of the treated states were early adopters of welfare reforms, then the present results would likely overestimate the effects of strong infertility treatment mandates on the mean age at first birth. In contrast, if early adopters constitute part of our synthetic group, then the estimated gap in mean age at first birth between treated and synthetic states would be underestimated. Reports show that early adopters [California, Colorado, Iowa, Michigan, Oregon, Wisconsin, and Utah (see Mead 2004)]—with the exception of California (which is neither a treated nor a control state)—are control states and, thus if anything, we should expect a downward bias in our estimates of the effects of the strong infertility treatment mandates on the average age at first birth.

## Appendix 2: Changes in infertility treatment laws

Four out of the six strongly treated states (Arkansas, Hawaii, Maryland and Massachusetts) revised their mandates during our sample period, i.e. before 2001. Table 12 briefly describes these revisions. The revisions in Arkansas and Maryland reduce coverage and hence would tend to decrease the estimated impact of the original mandates. The 1995 revision in Massachusetts established that the IVF procedures ICSI and ZIFT should be covered. ICSI is a particularly effective IVF procedure in cases of male infertility in which a single sperm is injected directly into an egg. Because this procedure was invented in 1991, it could not have been explicitly contemplated in the original mandate, although the original Massachusetts mandate covered IVF procedures. ICSI accounts for a large percentage of the fresh non-donor eggs or embryos, according to CDC (2001). The usage of ZIFT, however, has been declining gradually and in 2001 it accounted for less than 2 % of ART procedures (CDC 2001). Finally, the Hawaiian revision in 1995 is clearly an expansion of coverage to dependent unmarried individuals. Simple DID estimates of the effect of this revision on the marriage

**Table 12** Revisions of infertility treatment mandates

State	Original mandate	Revision	Description of revisions
Arkansas	1987	1991	Imposition of minimum and maximum benefits and setting standards
Hawaii	1987	1995	Dependents of the insured are also covered
Maryland	1985	1994	Exempt businesses $\leq 50$ employees from IVF coverage
		2000	Restricts coverage to 3 IVF attempts/live birth Exempts organizations with religious conflicts
Mass.	1987	1995	Extends coverage to ICSI and ZIFT <sup>a</sup>

Sources: Schmidt (2005b); [https://www.hrtools.com/http://us.firstvisitivf.org/display.asp?page=IVF\\_coverage\\_in\\_USA#state-law](https://www.hrtools.com/http://us.firstvisitivf.org/display.asp?page=IVF_coverage_in_USA#state-law)  
<http://www.resolve.org/http://www.fertilitylifelines.com/payingfortreatment/state-mandatedinsurancelist.jsp>

<sup>a</sup> For Massachusetts, see 1995-08 211 CMR 37.00, new infertility mandated benefits



probability of first-time mothers show either no effect or a positive effect for whites.<sup>43</sup> Hence, since there is no evidence that the Hawaiian revision decreased the marriage rate in the state, it is unlikely that it had a significant impact on the number of covered users.

## References

- Abadie A, Gardeazabal J (2003) The economic costs of conflict: a case study of the Basque Country. *Am Econ Rev* 93(1):112–132
- Abadie A, Diamond A, Hainmueller J (2010) Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *J Am Stat Assoc* 105:493–505
- Abramowitz J (2014) Turning back the ticking clock: the effect of increased affordability of assisted reproductive technology on women's marriage timing. *J Popul Econ* 27(2):603–633
- Benyamini Y (2003) Hope and fantasy among women coping with infertility and its treatments. In: Jacoby R, Keinan G (eds) *Between stress and hope: from a disease centered to a health-centered perspective*. Praeger series in health psychology. Praeger Publishers, Westport
- Bhattacharya J, Vogt W (2000) Could we tell if health insurance mandates cause unemployment? A note on the literature, Mimeo, New York
- Bitler M (2005) Effects of increased access to infertility treatment to infant and child health outcomes: evidence from health insurance mandates. In: *RAND labor and population*, WR-330
- Bitler M, Schmidt L (2012) Utilization of infertility treatments: the effects of insurance mandates. *Demography* 49(1):124–149
- Buckles K (2005) *Stopping the biological clock: infertility treatments and the career-family tradeoff*. BU dissertation
- Bundorf MK, Henne M, Baker L (2007) Mandated health insurance benefits and the utilization and outcomes of infertility treatments. In: *NBER working paper* 12820
- Centers for Disease Control and Prevention, American Society for Reproductive Medicine, Society for Assisted Reproductive Technology (2001) 2001 assisted reproductive technology success rates: national summary and fertility clinic reports. Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention (<http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5301a1.htm#tab4>)
- Centers for Disease Control and Prevention, American Society for Reproductive Medicine, Society for Assisted Reproductive Technology (2007) 2005 assisted reproductive technology success rates: national summary and fertility clinic reports. Centers for Disease Control and Prevention, Atlanta
- Collins J (2001) Cost-effectiveness of in vitro fertilization. *Semin Reprod Med* 19(3):279–289
- Friedberg L (1998) Did unilateral divorce raise divorce rates? Evidence from panel data. *Am Econ Rev* 88:608–627
- Grant J, Hoorens S, Gallo F, Cave J (2006) Should ART be part of a policy mix. Project Resource, RAND Europe
- Gruber J (2004) Is making divorce easier bad for children? *J Labor Econ* 22(4):799–833
- Gumus G, Lee J (2012) Alternative paths to parenthood: IVF or child adoption? *Econ Inq* 50(3):802–820
- Hamilton BH, McManus B (2012) The effects of insurance mandates on choices and outcomes in infertility treatment markets. *Health Econ* 21(8):994–1016
- Hao L, Cherlin AJ (2004) Welfare reform and teenage pregnancy, childbirth, and school dropout. *J Marriage Fam* 66:179–194
- Heckman J, Ichimura H, Todd P (1997) Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. *Rev Econ Stud* 64(4):605–654

<sup>43</sup> These DID estimates were obtained using the birth certificates data from 1987 (the date of the original mandate) to 2001, taking Hawaii as the treatment group against all non-treated states as controls. When controlling for race and education, we obtain essentially the same effect. We repeated the estimation for whites only and obtained a positive and significant effect of the 1995 revision on marriage rates of first-time mothers, which is in contrast to what would be expected.

- Heckman J, Ichimura H, Smith J, Todd P (1998) Characterizing selection bias using experimental data. *Econometrica* 66(5):1017–1098
- International Federation of Infertility Societies (IFFS) (2007) Surveillance 07. In: Jones HW Jr., Cohen J, Cooke I, Kempers R (eds) Supplement to fertility and sterility, vol 87, sup. 1
- King M, Steven R, Trent AJ, Flood S, Genadek K, Schroeder MB, Trampe B, Vick R (2010) Integrated public use microdata series, current population survey: version 3.0. [machine-readable database]. University of Minnesota, Minneapolis
- Kroeger S, La Mattina G (2015) Assisted reproductive technology and women’s choice to pursue professional careers. USF WP 0115. Mimeo, New York
- Lampi, E (2006) The personal and general risks of age-related female infertility: is there an optimistic bias or not? Working Paper in Economics No. 231, School of Business, Economics and Law, Göteborg University
- Machado MP, Sanz-de-Galdeano A (2011) Coverage of infertility treatment and fertility outcomes: do women catch up? In: CEPR DP 8445 and IZA DP 5783
- Mead LM (2004) State political culture and welfare reform. *Policy Stud J* 32(2):271–296
- Michalopoulos C, Bloom H, Hill C (2004) Can propensity score methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs? *Rev Econ Stat* 86(1):156–179
- Mookim PG, Ellis RP, Kahn-Lang A (2008) Infertility treatment, ART and IUI procedures and delivery outcomes: how important is selection?. Boston University, Mimeo
- Neumann PJ, Gharib SD, Weinstein MC (1994) The cost of a successful delivery with in vitro fertilization. *N Engl J Med* 331(4):239–243
- Ohinata A (2011) Did the US infertility insurance mandates affect the time of first birth? Center Discussion Paper 2011–102
- Schmidt L (2005a) Infertility insurance mandates and fertility. *Am Econ Rev* 95(2):204–208
- Schmidt L (2005b) Effects of infertility insurance mandates on fertility. In: Labor and demography 0511014 WPA
- Schmidt L (2007) Effects of infertility insurance mandates on fertility. *J Health Econ* 26:431–446
- Simon KI (2004) Adverse selection in health insurance markets? Evidence from state small-group health insurance reforms. *J Public Econ* 89(9–10):1865–1877
- Smith J, Todd P (2005) Does matching overcome Lalonde’s critique of nonexperimental estimations. *J Econ* 125(1):305–353
- Ward MP, Butz WP (1980) Completed fertility and its timing. *J Political Econ* 88(5):917–940
- Ziebe S, Devroey P (2008) Assisted reproductive technologies are an integrated part of national strategies addressing demographic and reproductive challenges. Human reproduction update. European Society of Human Reproduction and Embryology, Oxford University Press, Oxford , pp 1–10