Preliminary, comments welcome but please do not cite or quote without permission

Effects of Increased Access to Infertility Treatment on Infant and Child Health Outcomes: Evidence from Health Insurance Mandates[1]

Marianne P. Bitler
PPIC and RAND Corporation

First version: July 2002
This version: March 28, 2005
JEL classification: I3, JI
Key words: infertility, infant health, insurance mandates, twins, birth outcomes, fertility timing, impaired fecundity

[1]I thank Reagan Baughman, Julie DaVanzo, Chrissy Eibner, Susan Ettner, Jonah Gelbach, Deborah Haas-Wilson, Janice Hauge, Vivian Ho, Hilary Hoynes, Becky Kilburn, Jacob Klerman, Dave Loughran, Nicole Maestas, Barbara Schone, Kosali Simon, Lucie Schmidt, Jennifer Ward-Batts, Madeline Zavodny, Julie Zissimopoulos, and participants at the RAND Aging, Demography, and Population Seminar, the Melbourne Institute of Applied Economic and Social Research, the Economic Demography Workshop at the PAA, and the WEA for valuable comments. I thank Aaron Yelowitz for generously providing data on Medicaid expansions. I gratefully acknowledge the financial support of the National Institute of Child Health and Human Development (R03 HD046485), the National Institute on Aging, and the RAND Corporation. This work has not been formally reviewed or edited. The views and conclusions are those of the author and do not necessarily represent those of PPIC, the RAND Corporation or any other organization. Correspondence to Bitler at PPIC, 500 Washington Street, Suite 800, San Francisco, CA 94111 or bitler@ppic.org.

Abstract

Reproductive technologies have radically improved since the introduction of the first fertility drugs in the late 1960s. These technologies make conception possible for many couples who otherwise would have been unable to reproduce. Many of these technologies increase the probability of having a multiple birth, typically a more risky pregnancy. These technologies also provide women considering delaying reproduction with insurance against later infertility.

Live births resulting from use of advanced reproductive technology and other infertility treatments are a tiny share of all births (around 0.5% of all births in 1999 were due to use of advanced reproductive technologies), making it difficult to detect impacts of the advanced technologies. Detecting impacts of infertility treatments in general is difficult in public-use data because such data do not record use of such treatments. To detect impacts of the advanced technologies, we restrict ourselves to a sample—twin births—where births related to reproductive technologies are much more common. About 6% of twin births in 1999 were the product of in vitro fertilization or other advanced reproductive technologies. To examine impacts of infertility treatments more broadly, we also consider a sample of singletons.

We analyze the association between the dissemination of advanced reproductive technologies and other infertility treatments and infant and child outcomes during the 1980s and 1990s. We rely on variation in access to advanced reproductive technologies (ART) and other infertility treatments induced by state-level insurance mandates forcing insurers to cover or offer to cover use of infertility treatments, thus subsidizing use of these treatments.

We find significant impacts of the mandates on fertility and birth outcomes. First, we document that the mandates have affected fertility. Using the Detailed Natality data, we find that for mothers 30 and older, state infertility mandates are associated with a statistically significant 10 percent increase in the rate of live twin births, typically more dangerous pregnancies. Subsidized infertility treatment as proxied by the mandates is also associated with small but statistically significant negative effects on gestation, birth weight, and the 5-minute Apgar score for these older mothers. Effects for singletons are smaller in magnitude as would be expected given that use of infertility treatment in general and ART in particular is less common for mothers of singletons. We then examine whether there are longer-term impacts of this increased access to infertility treatments. Using data from the 2000 Census, we replicate our finding from the birth certificate data that twin births to older women are more common in states with mandates. Our Census findings are mixed concerning longer-term negative health outcomes for twins born to older mothers in mandate states. Overall, our findings for twin birth outcomes suggest that positive effects of investment by older mothers in their pregnancies are outweighed by negative impacts of either the infertility treatments themselves or by the selection into pregnancy of women with reduced fecundity.

1 Introduction

Reproductive technologies have improved drastically in the last 40 years. Beginning in the late 1960s when the first ovulation-inducing drugs were approved in the U.S., and continuing with the use of new technologies such as in vitro fertilization (IVF), these advances have enabled hundreds of thousands of pregnancies around the world. An implicit assumption behind the widespread use of these technologies is that children born to women using advanced reproductive technologies (ART) or other infertility treatments are similar to other children. In this research, we test this assumption by estimating the demographic and health effects of infertility treatments. Because we cannot simultaneously observe use of infertility treatments and birth and child health outcomes in standard data sets, we rely on variation in state-level mandates that require insurers to cover or offer to coffer infertility treatment. These mandates implicitly subsidize use of infertility treatment. We estimate reduced-form regressions relating fertility and infant and child health to the mandates.

This paper contributes to the literature on the effects of health insurance mandates. This paper also contributes to the ongoing debate about birth selection in health economics. At least three mechanisms could lead to children born via use of infertility treatment differing from other children. First, their parents might be different from other children's parents. For example, the probability that a woman will be infertile (raising demand for ART and other infertility treatment) increases with age. Many infertility treatments such as ART are also expensive, so parents using them are likely to be wealthier. Second, because these technologies are resource-intensive, in both time and money, these would-be parents are likely to value their children very highly and, thus, are also likely to take other steps to improve child health (such as obtaining high-quality prenatal care). Third, it is possible that some of the same medical issues that make it difficult for some women to conceive could directly affect the health of the infant—either through a genetic correlation between maternal and child health or through making it difficult for the mother to support the fetus.

These issues are of interest for two reasons. First, use of infertility treatments and ART is of interest to the medical community and to potential parents; thus, the demographic and health impacts of these technologies should also be of interest. Second, infertility treatments such as ART generate an archetypal birth-selection problem. As in the earlier literature on birth selection through the use of abortion (e.g., Gruber, Levine & Staiger (1999)), these infertility treatments affect the infant health distribution, but not by affecting the health of a fixed infant population. Instead these infertility treatments make it possible for a new subset of children to be born. Such birth selection raises important methodological and conceptual issues into which this work provides important insight.

Births after using ART are a tiny share of all births, so sample sizes from typical survey data sets are likely to be too small to detect direct impacts of access to ART on birth outcomes. However, many infertility treatments increase the probability of having a multiple birth—for example, use of IVF in 1998 was associated with a 30 percent chance of having twins compared to a chance of 2 percent without any infertility treatment. Induced ovulation exclusive of IVF accounted for 38 percent of higher-order multiple pregnancies (Martin & Park (1999)). Rates of multiple births have increased dramatically since introduction of these technologies; the twin rate has risen from 18.9 per 1000 live births in 1980 to over 27 per 1000 live births in 1997 (Martin & Park (1999)). Some of this increase is due to use of technology and some is due to more older mothers having children—with or without use of ART.

Our main source of identifying variation is state laws on insurance coverage of infertility treatment. A number of states passed laws in the late 1980s and early 1990s mandating that health insurers either cover or offer coverage of infertility treatment. These state mandates lower the price of obtaining infertility treatment for resident women seeking such treatment. By using variation from these mandates, we can differentiate between the impact of age versus that of gaining access to subsidized infertility treatment.

Using a subsample of singleton and twin births from the Detailed Natality data for 1981–1999, we find that, for older mothers, state mandates—which increase access to infertility treatments—are associated with a statistically significant increase of between 10 and 24 percent in the probability of having a twin birth, typically a more dangerous pregnancy than a singleton pregnancy.

Because births due to use of ART or other intense infertility treatments are rare in the population of singleton births, we focus primarily on health outcomes among twins. Births to women who use infertility treatment are a substantial share of twin births by the end of the sample period. Thus, we should be able to detect differences between twin birth outcomes for women using infertility treatment and other women, if such differences exist.

Using Detailed U.S. Natality data from 1981–99, we then estimate reduced-form regressions relating the state-level mandates to gestation, birth weight, and other outcomes for twin births and for singletons. Our results suggest that twins whose mothers live in states with mandates have statistically significantly shorter gestation durations and lower birth weights, and that the effects are larger for older women. We find smaller impacts for singletons as expected given the smaller share of singleton births that are due to ART or other infertility treatments.

Using pooled 1-percent and 5-percent Public-Use Microdata Sample (PUMS) data from the 2000 Census, we construct a panel of children born between 1982 and 1999. With these data, we replicate our finding that for older women, living in a mandate state the year before the child is born is associated with an increased probability of having a twin. Our findings concerning the longer-term health effects of having been born to an older mother in a mandate state are mixed.

The rest of the paper proceeds as follows. Section 2 discusses infertility and infertility treatment. Section 3 discusses related work and provides a theoretical motivation. Section 4 describes the Detailed Natality and PUMS data used here. The empirical models are presented in Section 5 and the main results and extensions in Section 6. Finally, Section 7 concludes.

2 Infertility and infertility treatment

Infertility is generally defined by clinicians as the inability to conceive after 12 months of unprotected intercourse (6 months for older women).[1] A related reproductive problem is impaired fecundity. Unlike infertility, which as commonly defined only applies to cohabiting and married women, any women could be considered as having impaired fecundity. A woman is considered to have impaired fecundity if she is not surgically sterile, and has had problems conceiving or carrying a pregnancy to term, or has been unable to conceive after a three year period of unprotected intercourse (Chandra & Stephen (1998)).

The three largest factors associated with infertility and impaired fecundity in women are problems in ovulation, blocked or scarred fallopian tubes, and endometriosis (the presence in the lower abdomen of uterine tissue). Sexually transmitted diseases are the most common preventable cause of damaged fallopian tubes. Among men, the most common causes of infertility are abnormal sperm or too few sperm. The Office of Technology Assessment (1988) reports that around one in five cases of infertility is not diagnosable.

Infertility and impaired fecundity are common. Abma, Chandra, Mosher, Peterson & Piccinino (1997) report that 2.1 million couples or 7.1% of married couples were infertile in the 1995 National Survey of Family Growth (NSFG) versus about 2.3 million couples in the 1988 NSFG and 2.5 million in the 1982 NSFG. While age-specific infertility rates have not increased over time according to the NSFG, the share of infertile couples trying to have a first birth rose dramatically between 1965 and 1988. Rates of impaired fecundity have risen considerably over the 1980s and 1990s. Chandra & Stephen (1998) report that rates of impaired fecundity rose from 8% of women 15-44 in 1982 to 13% in 1995.

Infertility treatment can be a long, arduous, and expensive process. The first stage of infertility

[1]Much of the information in this section is drawn from Office of Technology Assessment (1988). Women must also not be surgically sterile to be considered infertile.

treatment is a thorough examination of each partner's reproductive organs and their circulatory, endocrine, and necrologic functions. An early stage of treatment for women not ovulating normally is treatment with a fertility drug to stimulate ovulation. Drugs of various potencies are available, such as clomiphene citrate and gonadotropin-releasing hormones. The first of these drugs were approved for use in the U.S. in the late 1960s. Endometriosis can also often be treated with hormones.

More aggressive techniques such as in vitro fertilization (IVF), gamete intrafallopian transfer (GIFT), and zygote intrafallopian transfer (ZIFT) were pioneered in the early 1980s. In IVF, the oocytes (eggs) and sperm are combined in a laboratory. Once early embryos develop, they are transferred into the uterus. In GIFT, the oocytes are transferred immediately into the fallopian tubes and fertilization happens there. In ZIFT, the embryos are placed in the fallopian tubes. The first successful applications of IVF and GIFT in the U.S. led to babies born in 1981 and 1985, respectively. Currently, use of GIFT and ZIFT is relatively rare compared to use of IVF.

Another frequently used procedure that increases the likelihood of conceiving is intrauterine insemination (IUI). In this procedure, a woman is often given medication to stimulate multiple egg development and to help time when the egg(s) will be in the fallopian tubes. The sperm to be used is separated from the other components of semen in the laboratory, and then the concentrated sperm is placed in the cervix or high in the uterus using a catheter. Because women undergoing IUI are often superovulating, there is a non-trivial risk of multiple pregnancies with IUI.

Treatments for male infertility include artificial insemination and intracytoplasmic sperm injection (a form of micromanipulation, introduced in the 1990s, where sperm are injected directly into the eggs) as well as IUI. Surgery is also a treatment for some forms of male and female infertility.

Fertility services are used by a large and growing number of women. The 1995 round of the NSFG found that of 60.2 million women of reproductive age in 1995, 9.3 million or about 15% of women had ever sought treatment for infertility. The comparable figure in 1988 was 12% of women aged 15–44.

Abma et al. (1997) report that in 1995, 2% of women of reproductive age had an infertility visit in the last year, 3% had ever taken ovulation-inducing drugs, and 1% had ever used advanced reproductive technologies.

Infertility services can be quite expensive and are not covered by all insurance plans. A number of states have passed laws requiring insurance plans to cover or offer to cover infertility treatment (see Table 4 for a list of these states). The Office of Technology Assessment (1988) estimated that in 1988, a simple diagnosis of scanty or infrequent menstruation followed by treatment with fertility drugs would have cost around $3,668. At this time, around 30% of the couples would have successfully conceived via this treatment. A comprehensive evaluation for non-ovulatory causes of infertility would have cost a further $2,905. Tubal surgery to deal with blocked or damaged fallopian tubes would have cost around $7,118, and IVF around $9,376 for two cycles (two sets of fertilized eggs implanted in the woman). In 1988, Office of Technology Assessment (1988) estimated that around 70% of infertile women would have become pregnant by one of these treatments, under some assumptions about how many women would succeed at each stage of treatment. Neumann, Gharib & Weinstein (1994) calculate that the cost of a successful delivery via IVF ranged from $44,000 to $211,940 in 1992, depending on the cause of infertility, the age of the mother, and other factors. While these figures are dated, they show that infertility treatment can be quite costly. This suggests that having access to insurance coverage for infertility treatment might be quite valuable.

3 Literature Review and Theoretical Perspectives

First, we touch on the mostly descriptive medical and public health literature on the association between prior infertility or impaired fecundity and subsequent fertility and infant health. We only discuss the papers most relevant for this study, namely ones comparing the birth outcomes of women using infertility treatment with those of other women in a cross-sectional or case-control research design.

We note the difficulties associated with interpreting these previous findings as causal, referring to the existing economics literature on the production of infant health.

We then touch on existing work on health insurance mandates and their impacts on fertility and infant and child health. Finally, we present our theoretical framework and discuss the birth outcomes and child health outcomes that are the focus of the empirical work.

3.1 Infertility treatment, fertility behavior, and infant and child health

A number of papers in medical and public health journals examine trends in multiple births, in the use of infertility treatments, and in the outcomes of infants conceived via use of ART. The findings of the clinical literature on the association between infertility and birth outcomes are mixed. Some studies find that infertility is associated with having small-for-gestational age and preterm births while other studies find no such association. Many of the studies note one difficulty in identifying an association between infertility and birth outcomes. Infertile women who become pregnant are often older than other women, and age is independently associated with some negative birth outcomes.[2]

Schieve, Meikle, Ferre, Peterson, Jeng & Wilcox (2002) use registry data on all ART-assisted births for 1996 and 1997. They compare outcomes for infants conceived via ART with those for the full universe of infants born in 1997, finding that ART infants are of lower birth weight than the general infant population. Using Swedish data, Bergh, Ericson, Hillensjo, Nygren & Wennerholm (1999) find an increased risk of multiple births and an increased risk of prematurity and low birth weight for IVF singletons and for all IVF births relative to non-IVF births. They also find some increase in the frequency of congenital malformations among IVF singletons as compared to singleton births in the general population.[3] Dhont, Sutter, Martens & Bekaert (1999) find perinatal outcomes of singleton

[2]See Varma & Patel (1987), Sheiner, Shoham-Vardi, Hershkovitz, Katz & Mazor (2001), Levi, Raynault, Bergh, Drews, Miller & Scott (2001), and Tan, Doyle, Campbell, Beral, Rizk, Brinsden, Mason & Edwards (1992).

[3]Saunders, Spensley, Munro & Halasz (1996) use Australian data to match IVF infants with infants from the general population, finding similar results to Bergh et al. (1999).

pregnancies conceived via use of ART are significantly worse than those of singleton pregnancies conceived spontaneously. Zhang, Meikle, Grainger & Trumble (2002) use a cross-section of births from 1995–97, to examine the characteristics of women having multiple births. Compared to women having singletons, women having multiple births were more likely to be older, non-Hispanic white, more highly educated, married, and not to have any previous live births. These women had earlier and more frequent prenatal care. In this cross-section, Zhang et al. (2002) find that twins born to older mothers were at no higher risk than other twins, and triplets born to older mothers were at lower risk than other triplets, of being very premature, of being very low birth weight, or of dying in their first year. This pattern is reversed for singletons born to older mothers.

In the last few years, several studies have appeared that examine the longer-term health impacts of ART on children. Sutcliffe (2004) summarizes this literature, concluding that when born mature, ART children are as healthy as other children, once their higher probability of having congenital anomalies.

These last studies mentioned have the same goal as this paper, namely to identify the causal impacts of ART or infertility treatments more broadly. Clearly, the fact that mothers of multiples in the cross-section differed systematically along many observable dimensions from women having singletons raises questions about a causal interpretation of the Zhang et al. (2002) results. It is possible that the women of higher socioeconomic status are positively selected along some unobservable characteristics or alternatively have access to better medical care. Either way, there could be a positive association between pregnancy outcomes of these women and their unobservable characteristics that offsets a possible increase in risk associated with being older and/or using infertility treatments. Similarly, many of the other findings discussed are based on cross-sectional comparisons of ART infant and children with other infants and children, again raising questions of causality.

3.2 Birth outcomes and selection

We turn to a discussion of the economics literature on birth production functions, which focuses more attention on possible selection bias. Fertility behavior has been the focus of a considerable economics literature, beginning with Becker (1991). An extensive line of work looks at infant health in a production function context; this work is surveyed in Wolpin (1997). Beginning with Rosenzweig & Schultz (1983), papers have estimated birth outcome production functions and input and output demand functions for infant health. A frequent question of interest here is whether health inputs such as prenatal care are associated with any impact on infant health.

An important issue in evaluating the impacts of various inputs on infant health is that of selection. In this context, selection refers to the possibility that women who use a health input may differ systematically from other women in ways that are both unobservable to the analyst and unrelated to the use of the health input. Comparing outcomes for the two groups of women will show that they are systematically different, yet we know it is because of their unobservable differences and not because of differential use of the health input. There can be selection both in which women use health inputs and in the health distribution of infants who are actually born. Both types of selection can bias findings.

Some papers hypothesize that women who anticipate adverse birth outcomes because of information available to them (but not to the analyst) may be more likely to invest in these inputs, biasing down estimates of the positive effects of prenatal care, for example. An alternate hypothesis posits positive selection; women who initiate care early may engage in other forms of healthy behavior, leading researchers to overestimate the positive impacts of prenatal care.

Grossman & Joyce (1990) emphasize that the resolution of the pregnancy itself may also be characterized by selection (e.g., women who think they have very unhealthy fetuses may choose to abort these fetuses). Grossman and Joyce find evidence of positive selection among black but not white women choosing to carry their pregnancies to term. Donohue & Levitt (2001) finds that introduction of legalized abortion is associated with a decrease in crime rates for cohorts born after legalization although Joyce (2004) questions this finding, while Donohue & Levitt (2004) defends it. Gruber et al. (1999) show that average outcomes of infants born after the legalization of abortion were better than those of infants born before legalization, suggesting that access to legal abortion led to positive selection among children who were born. In a more recent paper, Ananat, Gruber & Levine (2004) suggest that this selection was due to a permanent and not transitory decline in births for some women.

This previous literature on birth selection suffers from an inability to separate out heterogeneity or selection among mothers from the impact of investment by mothers. Both heterogeneity and investment are expected to affect birth outcomes in the same direction. Our analysis does not suffer from this drawback. Users of ART and other infertility treatment are infertile and may have more difficulty supporting a fetus through an entire pregnancy than do other mothers despite having invested heavily in their pregnancies. The selection biases from these two factors (heterogeneity and investment) are expected to go in opposite directions.

Another literature, touched on above, documents that women in the U.S. are delaying childbearing, perhaps to accumulate more human capital or to advance in the labor market. This delay in fertility may bear costs if there is a substantial age-gradient in fertility, confounding the selection bias discussed above. Royer (2003) uses panel data to study the impact of age at birth on children's outcomes, finding that women over 35 and women under age 18 are more likely to have a preterm delivery than women 26-29, and that women over 35 also face a higher risk of infant death and of having infants with some abnormal condition.[4] It is important that we also control for this age gradient in order to identify a selection effect due to ART or other infertility treatments rather than confounding it with one due to delayed childbearing. We discuss this below in more detail, but point out that our use of variation

[4] Royer's estimates account for unobserved maternal differences using correlated random-effects. Notably, while her correlated random-effects estimates are quite different from the fixed-effects estimates, they are fairly close to her own adjusted cross-sectional estimates, which are similar to those from the medical literature (e.g., Cnattingius, Forman, Berendes & Isotalo (1992)).

induced by insurance mandates allows us to compare outcomes for older women in mandate states with that of other, observationally similar, older women.

3.3 Insurance mandates

Without asymmetric information, economic theory provides clear predictions for the impact of employer mandates on wages, employment, and insurance coverage; namely that they should all decrease (e.g., Summers (1989)). In such a setting, with homogeneous workers, wages should fall until (almost) all of the cost is covered by the worker's lower wages. With heterogeneous workers, firms may respond by cutting insurance availability or shifting employment. All of these predictions hold to some extent for the case of group-specific mandated benefits. However, in the presence of informational asymmetries, theoretical implications for the effect of health insurance mandates on coverage and on health outcomes are less clear. For example, with adverse selection, even if individuals value a particular form of coverage above than their own actuarially fair price, they may not be able to obtain coverage in the individual market without a mandate.

Practically, the Employment Retirement Income Security Act of 1974 exempts self-insured firms from state-level mandates, suggesting that imposition of mandates may encourage firms to self insure. To the extent that it is more expensive for small firms to self-insure, this suggests the costs of mandates may fall disproportionately on smaller firms (Jensen & Morrisey (1999)). A large share of workers are employed by self-insured companies, and presumably these firms are not affected by the mandates. However, depending on the correlation between demand for the service covered by the mandate and worker productivity, mandates could also compel self-insured firms to offer benefits they otherwise wouldn't in order to retain highly productive employees.[5] Theoretical ambiguities about the impacts of mandates suggest we turn to empirical evidence. However, the empirical literature on the effects

[5] Liu, Dow & Norton (2004) touch on possible spillovers of "drive through delivery" laws mandating minimum hospital stays for women post-delivery to women not directly covered by the laws.

of mandated health benefits on health insurance and health outcomes has mixed findings, with some papers finding little significant impact of mandates on health status or health care utilization (e.g., Klick & Markowitz (2003), Pacula & Sturm (2000), Bao & Sturm (2004)), while other papers find some significant impacts on health utilization and health status (e.g., Liu et al. (2004)).

There is little social science literature about how state-level mandates to cover new reproductive technologies have altered the price of having children or child outcomes for women with impaired fecundity.[6] Recently, several clinical papers have used cross-sectional data to examine the impact of insurance mandates on use of IVF and pregnancy outcomes. Massachusetts imposed a broad mandate to cover infertility treatment in 1987. Griffin & Panak (1998) use administrative insurance company data from Massachusetts for nine large insurance plans for 1986–1992. They find imposition of the mandate is associated with increased use of ART and a small increase in the per capita cost of providing infertility services for Blue Cross Blue Shield but a decrease in per capita cost for HMOs. Jain, Harlow & Hornstein (2002) compare aggregate numbers of IVF cycles and live births in 1998 for states with no coverage of IVF, partial coverage of IVF, and full coverage of IVF.[7] They find that women at clinics in full-coverage states underwent more cycles (per capita) than did women in partial-coverage or no-coverage states. However, women in full-coverage states also transferred fewer embryos, suggesting that insurance coverage of IVF might be associated with a decrease in the risk of high order births.[8] One problem with these studies using insurance coverage is their reliance on cross-sectional data.

Schmidt (2005) uses Detailed Natality data from 1981–2000 and differences-in-differences techniques to see whether first birth rates are higher in states with mandated coverage, controlling for age, race, state, and year. Schmidt (2005) finds that first birth rates are statistically significantly higher for

[6] Exceptions in the economics literature include papers by Schmidt (2005), Buckles (2004), Hamilton & McManus (2004*b*), and Hamilton & McManus (2004*a*), discussed below.

[7] Recall that one IVF cycle can involve implantation of one or, frequently, more than one fertilized embryo in the woman.

[8] Reynolds, Schieve, Jeng & Peterson (2003), using microdata for 1998, find that while women in states that mandate coverage of IVF indeed transfer fewer embryos, this only translates into a lower triplet or higher order birth rate for one state, Massachusetts.

women over 35 in states mandating coverage.[9] Buckles (2004) also uses the Detailed Natality data to look at age at first birth, finding mandates are associated with higher average age at first birth. She then looks at the impacts of mandates on labor force participation and wages, concluding that mandates lead to increased earlier and decreased later labor force participation and higher wages for all working women.

In a series of papers (e.g., Hamilton & McManus (2004*a*), Hamilton & McManus (2004*b*)), Barton Hamilton and Brian McManus create a theoretical model of the behavior of infertility clinics under competition and mandates and of the use of ART such as in vitro fertilization. They test their model using data from infertility clinics. Hamilton & McManus (2004*a*)) uses clinic data for 1995–2001 and 1987 on use of in vitro fertilization, the number of cycles undergone, the number of embryos transferred, the birth rate, and the rate of multiple births. Because their clinic data only contains information on cycles of ART such as in vitro fertilization, Hamilton and McManus differentiate between mandates that include IVF and are "universal" and other mandates. They find that universal insurance mandates increase cycles of IVF but cut the number of embryos transferred. That find that competition between clinics increases the number of ART cycles as well. Their clinic data do not allow them to tell whether more women are obtaining access to ART or whether women who already had access are undergoing more cycles of ART.

Hamilton and McManus's data do not contain information about use of other infertility treatments, thus it cannot capture all impacts of the insurance mandates unless the mandates had no effect on use of any infertility treatments but ART. Further, no states adopted mandates during their sample period, so they cannot include state fixed effects in their estimates. They argue that states with mandates are similar to states without mandates (and thus that there is no need to control for time-invariant

[9] Schmidt's data consist of birth rates within 5-year-age-group by race by state by year cells. Her findings are robust to controlling for cohort-year interactions, and account for a variety of other factors that are known to influence fertility behavior (e.g., abortion restrictions in the state).

differences across states). Hamilton and McManus support this contention by looking at treatment rates in 1987 as a function of the later mandates. Hamilton & McManus (2004*b*) looks at the impacts of competition on technology choice for clinics for 1987–2001, finding clinics in competitive markets were quicker to adopt new technologies.

3.4 Theoretical perspectives

Improvements in infertility treatment lower the cost to infertile women of having their own children rather than going without or adopting. This may increase fertility among couples who would have been unable to conceive before these new technologies were widely available. In effect, these technologies may have shrunk the tail of the fecundity distribution. If women at the lower tail of the fecundity distribution are more likely to have infants at the lower tail of the infant health distribution, all other things being equal, subsidized infertility treatment might be associated with worse infant and child health outcomes.

Alternatively, some types of infertility problems may have little or no correlation with infant health (e.g., blocked fallopian tubes). The existence of advanced reproductive technologies might also provide some women with insurance against being infertile, enabling them to delay childbearing. If there is an important age-gradient in birth outcomes, these women's infants may have worse health outcomes than if the women had not delayed childbearing. However, if these women accumulate sufficient human capital, or achieve higher socioeconomic status, they may invest more in the health of the infants, suggesting a possible offsetting effect of increased inputs on infant health. The overall effect of technology and its possible insurance value is theoretically unclear.

Any woman who goes through the more invasive medical interventions described above in Section 2 clearly wants any resulting children very much. Thus, one expects that women who respond to the insurance subsidies or who delay childbearing will invest more in fetal health during their pregnancy than will other women.

By using a pooled series of cross-sections of singletons and twins over the period 1981–9999, we can assess whether either the probability of having a twin birth or twin birth outcomes have changed over time as use of ART and other technology has spread. By also examining Census data on the prevalence of blindness, deafness, physical disabilities, and other conditions, we can see whether any health differences observed at birth have longer-term implications. By using data spanning a period before ART was used in the U.S., we ensure that the early period comparisons do not mix women who did and did not use ART. In the absence of both better technology for saving premature and otherwise at-risk infants and changes in the age structure of women giving birth, the pre-ART outcomes provide a benchmark against which to compare the later, partially ART-induced outcomes. We control for technology improvements in neonatal care by comparing outcomes for children of older versus younger mothers in each period; each group should have some access to neonatal care. We further control for improving neonatal care and other changing trends by including both year fixed effects and state-specific linear time trends. Finally, our differences-in-differences strategy uses state-level over-time variation in access to infertility treatments induced by infertility insurance mandates to compare women who have enhanced access to infertility treatments (because they live in states with mandates) to other women.

First, we use a random subsample of twins and singletons to show that the mandates have indeed resulted in changes in the twin birth rate. We then focus on birth outcomes—as birth weight, gestation, and Apgar score—that have been linked to later infant health. A large medical literature links low birth weight with adverse infant health outcomes (e.g., McCormick (1985) and Rees, Lederman & Kiely (1996)). A less settled literature examines the impacts of low birth weight on child morbidity (e.g., McCormick, Brooks-Gunn, Workman-Daniels, Turner & Peckham (1992)). Wolpin (1997) surveys the literature on the observed relation between birth weight and infant mortality. A series of studies by economists and others have found that birth weight is also correlated with later outcomes such as self-reported health status, earnings, and educational attainment (e.g., Behrman & Rosenzweig (2001) and Currie & Hyson (1998)). Almond, Chay & Lee (2004) discuss the existing literature on the importance of birth weight and offer evidence that the 5-minute Apgar score may be a better predictor of infant mortality among twins.[10]

There is also considerable evidence that premature infants are at risk of worse health outcomes (e.g., McCarton, Wallace, Divon & Vaughan (1996) and ACOG Committee on Practice Bulletins—Obstetrics (2002)). For example, nearly one-fifth of all very preterm births in 1999 did not survive the first year (Martin, Hamilton, Ventura, Menacker, & Park (2002)). The categories of low birth weight and very low birth weight, like the indicators for being premature, are taken from the International Classification of Diseases-Ninth Revision (ICD-9), and are key cutoffs in the birth weight and gestation distributions reported for births by the NCHS. "Extremely low" birth weight (being < 1000 grams) is a cutoff found by Almond, Chay & Lee (2002) to be correlated with a increase in infant mortality for twins.

Then, we use pooled 1-percent and 5-percent PUMS data from the 2000 Census to look at some longer-term child health outcomes. Our outcomes are measures of disability status asked about all persons 5 or older. The four measures we can look at are (1) does the person have long-lasting blindness, deafness, or a severe vision or hearing impairment?; (2) does the person have a long-lasting condition that substantially limits one or more basic physical activities?; (3) because of a physical, mental, or emotional condition lasting 6 months or more, does the person have difficulty learning, remembering, or concentrating?; and (4) because of a physical, mental, or emotional condition lasting 6 months or more, does the person have difficulty dressing, bathing, or getting around inside the home?[11]

[10]The 5-minute Apgar score is used to evaluate the condition of the newborn infant at 5 minutes after birth (Division of Vital Statistics, NCHS (2000)). It is a good predictor of surviving the first year. It is a summary measure of the infant's condition based on heart rate, respiratory effort, muscle tone, reflex irritability, and color. Each of these is given a score of 0, 1, or 2; their sum is the Apgar score. A higher score is better.

[11]We also plan to look at children's grade-for-age in future work but need to augment the 2000 PUMS data with data from the 1990 Census to do so.

These outcomes were selected for two reasons. The first is that these outcomes are ones the clinical literature associates with being born very preterm or of extremely low birth weight. For example, in a recent article, Vohr, Wright, Dusick, Mele, Verter, Steichen, Simon, Wilson, Boryles, Bauer, Delaney-Black, Yolton, Fleisher, Papile & Kaplan (2000) report neurodevelopmental, neurosensory, and functional outcomes of extremely low birth weight (401–1000 gram) survivors being tracked by the NICHD Neonatal Research Network. Extremely low birth weight children born in 1993 and 1994 were assessed at age 18–22 months. 51% of the children had normal neurodevelopmental and sensory assessments, and outcomes worsened as birth weight decreased. Colvin, McGuire & Fowlie (2004) review the evidence about neurodevelopmental outcomes after preterm birth, noting that very preterm infants (many of whom are also low birth weight) are at increased risk of vision and hearing problems, even though most preterm infants have good neurodevelopmental outcomes. The second and more practical reason for looking at these outcomes is that they are the only long-term health outcomes we are aware of available in the Census or any other large, public-use data set.

Our results suggest that being born in a mandate state to an older mother is associated with having a lower birth weight, a shorter gestation, and a lower Apgar score. This finding combined with the clinical evidence linking and extremely low birth weight births and developmental outcomes motivates our analysis of the association between being born in a mandate state to an older mother and these longer-term health outcomes.

Next, we turn to a more detailed discussion of the data we use.

4 Data

This paper uses two large data sets, pooled birth certificate data from 1981–1999 and pooled 1-percent and 5-percent PUMS data from the 2000 Census. First, we discuss the natality data and present basic summary statistics. We briefly discuss time series trends in the use of ART. We then describe and present summary statistics for the 2000 PUMS data.

4.1 Detailed Natality data

The data used in this paper are a subset of the National Center for Health Statistics Detailed Natality Data for the years 1981–99 (Division of Vital Statistics, NCHS (Various years)). The Detailed Natality data are either a one-half sample or the universe of birth certificates for live births in the U.S., depending on the state and year. The data include demographic information about the parents such as education, age, and marital status. For the analysis of birth rates, the data consist of 2-percent random sample of both singleton and twin births. For the analysis of infant health outcomes, the data includes all twin births in the Detailed Natality data or a 2-percent random sample of singleton births.[12] The data also include many pregnancy outcomes. For a large subset of the twin data, it is possible to probabilistically match the two twins in the pair as in Almond et al. (2002).[13] For matching records, we create indicators for whether the twins are both boys, both girls, or mixed-sex.

Mixed-sex twins must be dizygotic while same-sex twins can be either monozygotic (identical) or dizygotic (fraternal). Mixed-sex pairs and same-sex fraternal pairs will on average share 50% of their genetic material while, in the absence of mutations, same-sex identical twins will share 100% of their genetic material. Since twins that result from infertility treatment are less likely to be identical than are naturally occurring twins, an increase in the share of twins that are mixed sex suggests that more twins may be the result of infertility treatment. Many infertility treatments explicitly involve fertilizing and

[12] From 1985 through 1999, the Detailed Natality data include all births born to U.S. residents. Before 1985, some states submitted a 50% sample of birth certificates; for these years, the microdata contain a weight or one or two that allows calculation of statistics representative of the full population. Some states did not report maternal education or Hispanic ethnicity until the early 1990s; all regressions include dummy variables for these variables being missing or unreported.

[13] We do this by matching records for groups of births with the same state of occurrence, state of residence of the mother, county of residence of the mother, race of the mother, marital status of the mother, month of initiation of prenatal care of the mother, maternal education, place of birth of the mother, age of the father, month and year of birth, length of gestation, and number of prenatal care visits. Using this algorithm, 88% of the total records were matched to one other record, about 12% were matched to zero other records, and less than 0.05% were matched to more than one other record. Any match where one of the records had a weight of two (indicating it was from a state reporting only 50% of the birth certificates that year) was dropped (this resulted in dropping 1.3% of the matches).

implanting more than one egg in order to increase the chances of a successful pregnancy. Use of some ovulation-inducing drugs can cause more than one egg to be released, also leading to an increase in the chance of having a multiple birth (Callahan, Hall, Ettner, Christiansen, Greene & Crowley (1994)).

Figure 1 shows the average share of infants who are multiple births during 1981-1999, by state. Clearly, there has been an increase over this period in the share of live infants who are multiple births. The upward trend seems to pick up around the late 1980s. Figure 2 shows the average age of mothers of twins, by state, for the same period. This figure shows also that the average age of twin mothers has risen considerably over the period.

Average gestation and birth weight for twins have also gone down over time (not shown). This decrease in gestation and birth weight may partially be due to the dissemination of and tremendous improvements in technology for keeping infants alive over this period, and motivates inclusion of state-specific time trends in all regressions.[14] The composition of the pool of women giving birth to twins has also changed over this period (not shown). The share of twins that are mixed-sex has increased from 29% of twins in 1981–88 to 33% in 1995–99. This is consistent with an increase in the share of twins born to women treated for infertility as many treatments either encourage production of more than one egg or, like IVF, usually involve transferring more than one fertilized egg into a woman. The age of mothers has increased over the period, as has completed education, and the share of twins born to older women has also increased. Martin & Park (1999) note that more twins were born to women aged 45–49 in 1997 than in the entire decade of the 1980s.

Our empirical strategy relies on variation in state infertility mandates. Thus, it is interesting to see whether outcomes and demographics vary by whether the mother's state of residence had a mandate in effect the year before the birth.

Table **??** contains means for outcomes and some maternal and child demographics over two time

[14] For a discussion of the benefits of these new technologies, see Cutler & Meara (2000). For a discussion of factors affecting their dissemination, see Baker & Phibbs (2002).

periods: One when there was no state mandate for infertility treatment in effect during the year before the infant was born (column 1) and the other when there was such a mandate in effect (column 2). These means show that average gestation is slightly lower when mandates are in effect while birth weight is the same in both periods and Apgar scores are higher when mandates are in effect.

Table 2 shows average maternal age and the share of twins born to women in various age groups across the same two time periods. The share of twin births to older mothers is much higher in states with mandates in effect. The share of twin births to women 40 and older was 0.014 in no mandate state-years cells but 0.030 in mandate state-year cells.

Table **??** shows means of educational attainment among women of various age groups across the two time periods. Means and sample sizes are presented for records containing information about education and the number of records missing this data is also reported. Clearly, educational attainment of twin mothers has increased considerably, and the increase is larger (as a share of twin births in the age group) for older mothers.[15]

There has been a large shift in the distribution of women who had twins over the past 20 years. The means show that twin mothers look quite different depending on whether a mandate was present. These mean differences may be the result of the mandates or they may merely reflect changing trends. Next we discuss recent trends in ART use.

4.2 Use of advanced reproductive technologies

Beginning in the mid-1980s, the Society of Assisted Alternative Reproductive Technology of the American Fertility Society (later known as the Society for Assisted Reproductive Technology and the American Society for Reproductive Medicine, respectively) began a voluntary registry of clinic data on in vitro fertilizations, embryo transfers, and other ART treatments and their outcomes. In 1992, Congress

[15]The table does not report t-tests of the equality of means, but they are statistically significantly different between the two periods in this table for women 30 and older.

passed the Fertility Clinic Success Rate and Certification Act, which requires the CDC to publish clinic-specific success rates for ART procedures in the U.S. As a result, the total number of pregnancies and live births resulting from ART as well as the number and type of ART cycles are available from the registry and/or the CDC data. Unfortunately, registry data are not publicly availably at the state-by-year level. Data on the number of ART-associated deliveries are only available from 1989–1998. Figure 3 shows the total number of pregnancies, deliveries, and live births over the period. While there are no registry data for the period before 1985, it is unlikely that there were many ART-related pregnancies before that; the first live birth in the U.S. from an IVF procedure occurred in 1981. The first data reported by SART indicated that even in 1985, only 2389 IVF cycles, 26 frozen embryo transfers, and 56 GIFT procedures were performed, eventually resulting in just 160 live births (Medical Research International, American Fertility Society Special Interest Group (1988)). Figure 3 shows that the number of total pregnancies, deliveries, and live infants born after use of ART has increased considerably over the 1990s, again with much of the increase occurring after or around 1989, shortly after the first state-level laws would have impacted deliveries. Figure 3 also shows that many of the deliveries after 1989 were multiples (if all births were singletons, then the number of deliveries and live infants would be the same). This latter point is even more clear in Figure 4, which shows the number of infants per delivery and the number of pregnancies per delivery from 1989 on. The number of infants per delivery ranges from around 1.37 in 1989 to 1.45 from 1995–98. While unadjusted, the means do suggest that increases in the use of ART and multiple birth rate occurred around the same time as the first mandates were implemented.

4.3 PUMS data

The PUMS are 1- or 5-percent samples of data for occupied or vacant housing units, collected as part of the 2000 decennial Census. For persons in sampled units, the PUMS data include information on age, state of birth, race and ethnicity, disability status, educational attainment, and a host of other variables. We pool the two PUMS samples to increase sample size. By taking children aged 0–17 from the PUMS, one can create a panel of children still alive in 2000 who were born from 1982–1999. The only children omitted from this sample are children with no permanent home, children residing in group quarters, children who have died between birth and the 2000 Census, and children who have left the U.S. between birth and the 2000 Census. For children in the primary family, we also know their relationship to the householder and can infer whether their mother is in the household. By matching children of the same age and the same relationship to the householder whose age and relationship variables were not allocated, one can identify multiple births. For example, two children would be identified as twins if they were the only two children in the primary family of the same age and the and same relationship to the householder. Because the 2000 PUMS does not contain quarter of birth, this procedure will incorrectly identify some sets of siblings who were born only 9–11 months apart as multiple births. However, this measurement error should mostly attenuate our coefficients rather than introducing systematic bias.

By restricting the sample to grandchildren or natural-born children of the householder, we can be more sure these "twins" are primarily biological siblings. We can also link the children's records to information about their "mother." If the child is a biological child of the householder, the "mother" will be the biological mother if she is the householder and the wife of the biological father otherwise. If the child is a grandchild, we can use the subfamily number and subfamily relationship to tie children to their likely "mothers."[16]

For these grandchildren and biological children of the householder, we can estimate the mother's age at the child's birth as the mother's current age minus the child's age. Unfortunately, the only

[16]Subfamily information is only available for related subfamilies (married couples with or without never-married minor children or single parents with never-married minor children), where some subfamily member is related to the householder. For subfamily units where no member is related to the householder, no relationship information is available. Thus we cannot match children to their parents in such cases.

educational attainment measure we have for the mothers is current education, which is not necessarily the same as the mothers' education at the time of the birth(s). We can then link the child's record to characteristics of the state where the child was born in by state of birth and the approximate year the child was conceived. Year of conception is defined as $2000 - age - 1$. Because the disability measures are only reported for children aged 5 and up, we restrict the PUMS sample to children for whom the disability measures were reported. Table **??** contains means for children 5–17 who are singletons or twins, who are grandchildren or biological children of the householder, whose records have no allocated data on age or householder relationship, and whose "mother" is in the household. Column 1 contains means for the probability of being a twin birth and mother's age at birth for children born in states with no infertility treatment mandate the year they were conceived and column 2 means for children born in states with mandates while they were conceived.[17] We see that in the PUMS data, as in the Detailed Natality data, twin births are more likely in mandate states and mothers are generally older at birth in mandate states. Table 6 contains summary statistics for the disability measures, for the twin sample only, by mandate status during the year of conception. Here we see that impairments appear more common for twins born in states without mandates in place at conception.[18]

Next we turn to a discussion of our empirical strategy.

5 Empirical Model

This paper analyzes the impact of infertility treatments on fertility behavior and health outcomes. One naive approach would be to directly regress birth outcomes on indicators for use of infertility treatment. In fact, as discussed above, there are several papers in the medical literature that compare

[17] There will be some miscoding of state of conception here, as with the birth certificate data, but only if families move between conception and birth. This is likely not very common.

[18] Also note the rate of twinning in the PUMS data is slightly higher than that found in the Natality data. For example, the pre-mandate average rate of twinning was approximately 0.023 versus the 0.031 reported here. This likely is partially due to our miscoding some siblings as twins but could also be due to compositional differences between twins still living with their parents in 2000 and all twins.

birth outcomes of women using ART to those of other women giving birth. An issue with these case-control studies is that couples using ART are likely to be a selected sample. These couples are the parents who most want to have children, thus they may invest more in their children. Obviously, these couples are also at the lower end of the fecundity distribution, which may suggest that their infants have worse health endowments. A further issue is that, in addition to being potentially endogenous, use of infertility treatment is rarely recorded in public-use data sets. If it were, we could use the mandates as candidate instruments.

In this paper, we compare women who are likely to have easy access to infertility treatment to women who do not. Infants born via use of ART or other infertility treatments, while growing in number over the 1980s and 1990s, still represent a small share of total births. We focus on a population where they are likely to be a larger share of overall births, namely twins. We further focus on older women as the group most likely to be using these treatments. Society for Assisted Reproductive Technology, American Society for Reproductive Medicine (2002) reports that 52% of IVF cycles in 1998 were for women 35 or older. A joint report by National Center for Chronic Disease Prevention and Health Promotion, ASRM, SART, RESOLVE (2002) notes that 70% of cycles in 1998 were for women 30–39. Putting these two sets of numbers together suggests that about 81% of IVF cycles in 1998 were undergone by women 30 or older.

The econometric approach we take is to run least squares regressions of outcome measures on demographic covariates, state-level controls, state and year fixed effects, state-specific time trends, policy variables related to infertility mandates, and interactions of the policy variables with maternal age indicators.[19] The regressions have the following form:

$$y_{ist} = X_{ist}\delta + S_{st}\alpha + IT_{st}\beta_1 + A_{ist}ITst\beta_2 + \theta_s t + \gamma_s + \nu_t + \epsilon_{ist}. \quad (1)$$

[19]Results for outcomes that are binary variables are robust to running the regressions as probits.

Here, y_{ist} is any of a series of outcomes for an infant or child. For the fertility regressions, y_{ist} is simply an indicator for being a twin in a pooled sample of twins and singletons from the Natality data or from the PUMS. Health outcomes from the Detailed Natality data include length of gestation; indicators for the birth being premature (< 37 weeks) or very premature (< 32 weeks); birth weight; indicators for the birth weight being low (< 2500 grams), very low (< 1500 grams), or being extremely low birth weight (< 1000 grams); and the 5-minute Apgar score. For each birth outcome in the Natality data, we only use the balanced panel of states and years where the outcome was reported by the state for every year. The disadvantage of this sample selection is that regressions for different outcomes may use different samples but the advantage is that it maximizes the sample for each outcome. We investigate the sensitivity of the results to this sample restriction in Section 6.5 below.[20] These outcomes are ones the literature associates with infant mortality and later worse infant outcomes.

In the PUMS data, y_{ist} is any of the four disability measures: Being blind or deaf, or having a severe vision/hearing impairment; having a long-lasting condition that substantially limits physical activity; having a condition lasting at least 6 months that causes difficulty learning, remembering, or concentrating; and having a condition lasting at least 6 months that causes difficulty dressing, bathing, or getting around inside the house. These are also measures the medical literature finds more prevalent among infants born very prematurely or of extremely low birth weight.

X_{ist} is a vector of demographic characteristics, including controls for the mother's age group (20–24, 25–29, 30–34, 35–39, 40–44, and 45 and older), race (black, Asian or Pacific Islander, American Indian, and other non-white), Hispanic ethnicity (or an indicator for Hispanic ethnicity not being reported), and education (high-school graduate, some college, four-year college degree, or missing). Regressions using the Detailed Natality data also control for the number of previous live births (1, 2, 3, 4, 5 or more,

[20] These restrictions are more onerous for some outcomes than others. New Mexico is the only state excluded from the gestation regressions. California, Delaware, and Oklahoma are excluded from the Apgar score regressions.

or missing).[21] We also control for the infant's or child's sex. Specifications run on the pooled, matched-twin data also contain indicators for the twins being mixed-sex or for both being girls. Educational attainment is included as a proxy for permanent income or resources. The other variables are included because of the evidence that birth outcomes vary by race, ethnicity, age of the mother, gender of the infant, and type of multiple birth (identical or not).

S_{st} is a vector of state-level demographic, labor market, public assistance, and, in some specifications, health care access variables. These variables include the unemployment and aggregate employment growth rates, real median income for a family of four in the state, real annual AFDC/TANF benefits for a family of 4, the cutoff for Medicaid eligibility for a pregnant woman in the state as a share of the federal poverty level, the share of the state population that is under the poverty level, the share of births to unmarried mothers in the state, and the percent of the total state population that is black or Hispanic. These control for state characteristics that may be associated with fertility or the level of health care available to women. In some specifications discussed in Section 6.5, S_{st} also includes variables that control for the level of neonatal care technology available in the state. Means for these state-level variables for the sample of twins from the Natality data are presented in Appendix Table 7.[22]

The γ_s and ν_t terms represent state and year fixed effects. The state (year) fixed effects control for unobserved factors that differ across states and not over time (over time and not across states). Unobservable determinants are captured by ϵ_{ist}. The main specifications also include state-specific linear time-trends, $\theta_s t$. In the fertility regressions, these trends may capture changes in social norms. In the infant health regressions, the trends help capture dissemination of advanced technologies to

[21] We do not include number of siblings in the PUMS regressions because we only know about their existence if they are still living at home. This would be more of a problem for the older children and possibly for the children of older mothers, who may have had other children while young.

[22] Sources for all state-level variables are discussed in a Data Appendix, available upon request.

prolong the life of high risk infants.[23] All regressions and summary statistics are weighted. For the birth certificate data, weighting is necessary because in the early 1980s, some states only provided a 50% sample of birth certificates. For the PUMS data, use of weights accounts for varying sampling probabilities and non-response adjustments.

All standard errors are clustered at the state-by-year level. This accounts for the fact that the key policy—infertility mandates—only varies at the state-by-year level. despite policy variation only Bertrand, Duflo & Mullainathan (2004) and Kezdi (2002) raise concerns about serial correlation with differences-in-differences methodology using state policy reforms, particularly when the reforms stay on once implemented. The statistical significance level of key coefficients of interest was substantively the same with state-level clustering as with state-by-year level clustering.

In order to interpret the coefficients on the age by infertility mandate interactions as estimates of the impact of being older in mandate states on infant health outcomes, we must rule out one alternative. All of the mandates occur in the later part of the sample and no mandate states have subsequently lifted their mandates. If older women in the mandate states are disproportionately receiving care at hospitals with more advanced technologies for saving babies as compared to younger women and older women in the earlier periods, then the older women's more sickly fetuses may be more likely to survive, leading to a spurious association between being an older mother in a mandate state and adverse infant health outcomes. However, the younger women in the mandate states will also presumably have access to these technologies. As noted above, all infant health regressions include state-specific time trends as a imperfect proxy for dissemination of these baby-saving technologies. For the alternative hypothesis to bias the results, these new technologies would also have to be more easily available to older women than to younger women in mandate states post-mandate as compared to time periods with no mandates.

[23] One concern with this strategy is that trends may not adequately capture dissemination of these technologies. However, these advanced technologies may themselves spread because they are demanded by women who are delaying childbearing. Section 6.5 investigates the sensitivity of the estimates to including controls for neonatal intensive care and intermediate care availability within the state.

Simple cross-sectional differences in access will be captured by the main age effects.

Our main focus is on the coefficients of IT_{st}, a vector of dummy variables for state-level laws mandating insurance coverage of infertility treatment, and interactions of IT_{st} with a dummy for the mother being 30 and older or 35 and older. We interact insurance status with age because older women are those most likely to be suffering from infertility or impaired fecundity, and thus the interaction allows the possibility of a larger impact for older women. Alternatively, one could view the younger women as a control group for the older women.

If a state has passed a law mandating that health insurers cover or offer to cover infertility treatment, then the state is forcing insurers to subsidize infertility treatment. In themselves, these mandates do not automatically represent a subsidy. If the insurance market is working perfectly and there are no information problems or other inefficiencies, absence of such coverage would merely imply people were unwilling to pay for such coverage. However, even in an efficient market, if such mandates forced insurers to pool infertility treatment with other treatments, it could result in infertility treatment being subsidized. In the case of adverse selection where insurers cannot identify who likely users of infertility treatment are, mandates may also have real effects.

The coefficients on the insurance variables represent the impact of these subsidies on infant and child outcomes. There are three sets of insurance variables—we focus on the first two. The first set is simply an indicator for whether the state has any mandates concerning infertility treatment and its interaction with maternal age. The second set splits this law indicator into two variables, an indicator for whether the state has a law mandating that health insurers cover or offer to cover infertility treatment specifically excluding IVF and an indicator for whether the state has passed a law mandating that insurers cover or offer to cover such treatment including IVF. The third set splits the "any law" indicator by whether the mandate forces insurance companies to cover treatment or merely offer coverage to employers purchasing insurance policies.

Table 4 contains a list of states with mandates and indicates the year the state passed the law, whether the law mandates coverage or that the insurer offer coverage, whether the mandate specifically excludes IVF or does not, whether the mandate requires insurers to cover or offer coverage, and how the mandate applies to HMOs.[24] The indicator variables are set to one for every year after the year the law passed (treatment and a successful pregnancy usually take at least 9 months). Robustness regressions also include two-year lags of the laws.

There is considerable cross-state over-time variation in when and where the laws were passed during 1981–99 (statistics not in tables). 3% of the twins in the Natality data were born to mothers living in states where a mandate does not exclude IVF from 1981–88, while 17% of twins born in 1989–94 and 20% of the twins born in 1995–99 were born to mothers living in such states. A similar proportion of mothers of twins lived in states with laws mandating insurance coverage but excluding IVF.

The age-by-insurance status variables represent the differences-in-differences impact on birth outcomes for older women in mandate states periods versus older women in non-mandate states as compared to younger women in both sets of states.

We now turn to the results.

6 Results

Section 6.1 discusses the results of regressions predicting the probability of having a twin versus a singleton in the Detailed Natality and PUMS data. Section 6.2 then discusses the results for regressions predicting health outcomes for twins in the Natality data, Section 6.3 the results for health outcomes for singletons in the Natality data, and Section 6.4 the results for twins in the PUMS. Finally, Section 6.5 discusses various robustness tests.

[24] Information about the state laws was taken from National Conference of State Legislatures (2002) and cross-checked with law information available at `http://www.asrm.org/Patients/insur.html`. These laws do not cover self-insured companies' health insurance plans; as this is preempted by ERISA.

6.1 Mandates and the probability of twinning

Table 7 reports coefficients on our key mandate variables and their interactions with the mother being at least 30 for regressions using a 2-percent random sample of singleton and twin births from the 1981–1999 Detailed Natality data. Each column presents coefficients from a different regression including a different set of mandate variables. The first column suggests that twin births are more likely for women over 30 in mandate states. The coefficient of 0.003 represents about a 10 percent increase in the probability of having a twin birth, compared to the pre-mandate mean of 0.0295 for women over 30. Turning to column 2, we see that whether or not the state includes IVF in the mandate, women over 30 are more likely to have twins, and the effects are similar in size for the two kinds of mandates. The effects range from about a 10 percent to about an 11 percent increase relative to the pre-reform mean. Finally, column 3 presents results with the mandate variable allowed to vary by whether it is a mandate to cover or offer to cover treatment. As one might expect, the age-interaction with a mandate to cover has a large, statistically significant coefficient.

Table 8 presents results from regressions similar to those reported in Table 7 but with the mandate variables interacted with the mother's age being at least 35. Again, as expected, the effect for these older women is larger in magnitude, although the signs and pattern of significance are similar to those in Table 7. One exception is that for women 35 or older, even mandates that only require insurers to offer infertility treatment are associated with a statistically significant increase in the probability of twinning.

Finally, Table 9 shows results from regressions predicting twin births among children 5–17 using PUMS data. Here, the mandate variables are interacted with mother's age being at least 30. The "any law" variable interacted with mother's age being at least 30 is positive but no longer significant (although it is a statistically significant 0.002 (p-value of 0.030) if the sample is expanded to include children aged 0–4). Column 2 shows that for the children 5–17 in the PUMS, only women 30 and older living in states with mandates which exclude IVF face an increase in twin birth rates (this same coefficient is a larger 0.0023 (p-value of 0.114) if the sample also includes children 0–4). These coefficients are smaller in magnitude than we obtained with the birth certificate data, as one might expect given likely attenuation bias resulting from measurement error in identifying the twins. Appendix Table 3 contains the analogous results with interactions with the mother being 35 or older. Here, the age interaction for "any law" is statistically significant at the 10 percent level.

We next turn to the results for impacts of the mandates on infant health.

6.2 Mandates and infant health outcomes for twins: Detailed Natality data

The infant-health results for twins are reported in Tables 10–17 for the Detailed Natality data. Each table has a similar format and contains selected coefficients from two separate regressions predicting each outcome.[25]

Table 10 contains selected coefficients from two regressions predicting length of gestation; each of the other tables contain coefficients for two regressions predicting other outcomes. The only coefficients shown are those on the main insurance variables and the interaction of the insurance variables with being 30 or older. The first column presents results from regressions with dummy variable for the state having any mandate and its interaction with the mother being older than 30 and the second column presents results from regressions with two separate dummy variables for the type of insurance mandate the state has (including or excluding IVF) and their interactions with the mother being at least 30.

Tables 11 and table:regsoutanylaw-ivf:vpremmi contain results from regressions predicting being a premature birth and being a very premature birth. We focus on the gestation results but include the other specifications because they examine points in the gestation distribution other than the mean.

[25]Coefficients on the other control variables are presented for the regressions predicting length of gestation in Appendix Table 2. Later drafts will also present results including mandate variables that vary by whether they require insurers to cover or offer to cover treatment.

Column 1 of Table 10 shows that living in a state with any infertility mandate is associated with a statistically significant but small decrease in gestation of around 0.13 weeks, and that this effect is 0.05 weeks larger in magnitude for women over 30 (the mean at baseline is 35.91 weeks). Column 2 of Table 10 shows that forcing insurers to cover infertility treatment but not IVF is negatively associated with length of gestation. The coefficient suggests a decrease of around 0.17 weeks on a baseline of 36.3 weeks, and is larger in magnitude for the older women who face a 0.09 week shorter pregnancy than other women in the no-IVF mandate states. This is not a large effect. However, the only women who should be impacted by this law are those using infertility treatment, thus a zero impact on other women is being averaged with a larger effect on these women, leading to a small overall effect.

A simple back of the envelope calculation gives a sense of the possible overall magnitude of these results for the "treated" group. Society for Assisted Reproductive Technology, American Society for Reproductive Medicine (1999) suggests that around 6360 twins were born after use of IVF, ZIFT, GIFT, frozen embryos, or a host uterus from pregnancies initiated during 1996. Suppose twice as many women had twins due to any use of infertility treatment as did due to these treatments and that two thirds of the ART using women are at least 30 (as opposed to about 46% of all twin mothers). Suppose further that all twin births to women 30 and older were evenly split between insurance-mandate states and other states. Then, if we treat the younger women as a control group for the older women, the effect on gestation of the state mandate in states with mandates for women 30 and older using ART would be around $-(0.085 * 104,000 * 0.461 * 0.5)/(6360 * 2 * 0.67 * 0.5) = -0.48$ weeks, if all the twins were born in 1997 (104,000 twins were born in 1997). This assumes the mandates had no effect on women not using ART and no effect on women under 30. This may be an overestimate of the effect as the share of twins born after use of infertility treatment as a fraction of total twins should be larger in states with insurance mandates than in the full twin population. This would reduce the factor inflating the population effect to the effect on the "treated" group.

Results in Table 11 and 12 for regressions predicting premature and very premature births suggest that the impact of the health insurance mandate variables is larger for older women in states with mandates that exclude IVF. Older women in these mandate-but-no-IVF states are significantly more likely to have a premature birth (the coefficient is 0.012 (p-value of 0.000) on a baseline of around 0.48). Effects on being very premature are also significant for the older women interacted with "any law" and older women interacted with mandate-but-no-IVF variables.

Tables 13–16 present results for regressions explaining birth weight, low birth weight, very low birth weight, and "extremely low" birth weight (being under 1000 grams). Column 1 of Table 13 shows that living in a state with any infertility mandate is associated with having a 15 gram lighter infant. For women 30 and older, mandates are associated with having a further 17 grams lighter infant. Column 2 shows that the impact is again significant only for women in mandate-but-no-IVF states, although the point estimate for living in a mandate-and-IVF-covered state is also negative. For women of at least 30, living in a state with a mandate excluding IVF is statistically significantly associated with having a 24 gram lighter baby.

The pattern of the coefficients is similar for regressions predicting low birth weight, very low birth weight, or extremely low birth weight (Tables 14–16). The mandates are associated with an increase in having lighter infants for older women, and the effects are concentrated in the mandate-but-no-ivf states.

Table 17 presents results for regressions predicting the 5-minute Apgar score. Here the story is similar, namely living in a state that mandates coverage but excludes IVF is associated with having an infant with a lower Apgar score. However, women 30 and older in states with mandates that include IVF also have infants with lower Apgar scores (and the coefficient of -0.034 is larger in magnitude (more negative) for the mandate-with-IVF interaction than the coefficient of -0.022 for the mandate-but-no-IVF interaction). The baseline mean 5-minute Apgar for twins is 8.5.

Results for regressions interacting the mandate variables with the mother's age being at least 35 are generally similar to those discussed here (not shown). Effects of the interactions tend to be larger in magnitude when significant, and the significant results are still mainly for the mandate-but-no-IVF variables.

In sum, the results for the period 1981–99 Detailed Natality data suggest a statistically significant but small negative association between important health outcomes—gestation, birth weight, and the 5-minute Apgar score—and being an older mother living in a state that mandates insurance coverage of infertility treatment but excludes IVF. They also suggest that all women giving birth in states with any mandate had worse birth somewhat outcomes, although this could merely be picking up changing trends. If one interprets the younger women as a control group for the older women, the age-interactions are differences-in-differences in differences estimates of the impacts of mandates.

One might expect the effects of insurance mandates to be larger in states with mandates covering IVF than in states with mandates excluding IVF. However, there are alternate hypotheses that are consistent with both our findings and with the findings from the medical literature. If women in mandate-but-no-IVF states substitute away from IVF and toward intrauterine insemination or other procedures that increase the probability of multiple births, then there may be more infertility treatment related twins born in the mandate-but-no-IVF states than in the mandate-and-IVF-covered states.[26] Alternatively, Jain et al. (2002) suggest that women in mandate-but-no-IVF states may be transferring more embryos since they may be able to afford fewer cycles of IVF than women in mandate-and-IVF-covered states. Regardless, these findings suggest that insurance coverage of infertility treatments may have unexpected consequences.

We anticipated finding impacts of the mandates among twins if there indeed were any impacts of the mandates. Still, one might expect the twins results to reflect predominately use of ART. It is

[26] We thank Vivian Ho for suggesting this hypothesis.

possible that findings for singletons would instead pick up use of less-aggressive infertility treatments. The magnitude of the estimates should also be smaller for singletons, given the smaller share of women in mandate states actually using infertility treatment, so looking at singletons should provide a specification check on our twin results. We present results for the 2-percent sample of singleton births in the next section.

6.3 Mandates and infant health outcomes for singletons: Detailed Natality data

Table 18–20 present results for the 2-percent sample of singletons from 1981–1999 for regressions predicting gestation, birth weight and Apgar score. The signs of the mandate-age interactions are similar to those from the twin regressions, and many are still statistically significant. Further, the magnitudes of the coefficients is considerably smaller. Here the negative effects for women over 30 for gestation and Apgar score are concentrated in the mandate-and-IVF-covered states. The negative effects for birth weight are only statistically significant for older mothers in the mandate-but-no-IVF states. Many of the main effects of mandates are also in the direction of worse infant health and are statistically significant.

Next, we examine the longer-term health impacts of these mandates.

6.4 Mandates and child health outcomes for twins: PUMS data

Tables 21–24 present results from regressions predicting longer-term health impacts for twins born from 1982–1995. Here the mandates are interacted with an indicator for the mother being at least 30. While most of the point estimates for the mandate-age interactions are positive (being born in a mandate state to an older mother is associated with a higher probability of having the disabilities), none of the coefficients are significant at standard levels. Results from regressions estimated with mandate age interactions for the mother being at least 35 are similar, although more of the coefficients approach significance. For example, the coefficient on "any law" interacted with mother 35 and older is 0.010 (p-value of 0.111) for the regression predicting physical disability. That for the interaction of mandate-but-no-IVF with the mother being at least 35 is 0.016 (p-value of 0.082) for the regression predicting having a condition that causes difficulty learning, remembering, or concentrating. These findings are preliminary. We hope to add ACS data from 2001–2003 to see if the coefficients remain insignificant. We also plan to add 1990 Census data so we can see if grade-for-age is affected by the mandates.

6.5 Robustness

We experimented with a number of subsamples and specifications to test the robustness of the Natality results, and plan to also test the PUMS results. Below, we discuss the impact of these changes on the main results for gestation, birth weight, and the 5-minute Apgar score.

In general, restricting the Detailed Natality samples to women with no previous live deliveries before this pregnancy, to college graduates with no previous live deliveries before this pregnancy, to twins who could be matched, and to more homogeneous samples of matched twins did not substantively change key coefficients. Including single year of age dummy variables did not change key coefficients of interest nor did estimating regressions with the unbalanced panel of all observations available for each outcome.[27] Results from regressions including measures of neonatal care technology dissemination are also quite similar to the main results.

Women who are only able to carry a pregnancy to term due to new technologies may be at the highest risk for poor pregnancy outcomes. Thus, we expect the magnitude of the coefficients on the insurance variables and their interactions with age to be larger, and the effects more negative, when the sample is restricted to women with no previous live births before this pregnancy. This holds for regressions explaining gestation, birth weight, and the 5-minute Apgar score. For example, in the gestation

[27]Results discussed in this section are not shown. All are available from the author on request.

regressions, the coefficient on the indicator for a state law mandating coverage of infertility treatment but excluding IVF increases slightly when the sample is restricted to nulliparous women, although the coefficients are not statistically significantly different from one another.Regressions explaining birth weight and the 5-minute Apgar score show a similar pattern when the sample is restricted to women with no previous live births before this pregnancy.

One possible concern with the identification strategy is that without detailed information about where these women gave birth and what neonatal-care technologies their health care providers had access to, there may be negative coefficients on the age-by-mandate interactions because the older women are wealthier and give birth at hospitals that have better doctors and more advanced equipment. Thus, very sickly fetuses may be more likely to be born alive for these women than for younger, less wealthy women (although this concern may be more appropriate for outcomes such as neonatal mortality). This criticism could also apply to the estimates of the impact of the insurance laws without age interactions if access to better technology were confined to women in the mandate states. Regardless, this bias should be reduced or eliminated because we include state-specific linear time trends, a national time trend, and individual characteristics such as race and education. Another way to address this concern is to restrict the sample to women who have completed four-year college degrees and have no previous live births before this pregnancy (are nulliparous). These women should be equally likely to have access to the highest quality medical care (for each year) over the period, and the time trends should pick up any overall technology improvements. In this sample of four-year college graduates with no previous live births before this pregnancy, coefficients on the key variables of interest are generally similar in sign and magnitude to those in regressions that include all women with no previous live births before this pregnancy.

Next, we restricted the sample to the set of all births where the twin pairs could be matched. In addition to the other controls, these regressions also include indicators for the twins being same-sex female twins or mixed-sex twins. Again, the results are quite similar to the main results for the birth outcomes. Further restricting this matched-twin sample to three subsamples—mixed-sex twins, same-sex male twins, and same-sex female twins—yielded no surprises.

Dissemination of advanced neonatal care technology may reflect demand for these technologies from women seeking infertility treatment. While it is interesting to see how inclusion of measures of these technologies affect the results, because they may be demand-induced, their coefficients may be biased. Nonetheless, we pooled information from the American Hospital Association surveys for 1982, 1984, 1991, 1992, 1994, 1995, and 1997 on the number of obstetric beds, neonatal intensive care beds, and neonatal intermediate care beds in each state. We created variables measuring the total number of available hospital beds of each type per 1000 women aged 15–44 in each state. For years with missing data, we used the values for the previous year, and for 1981, we used values from 1982. Including these measures in the regressions hardly changed the key coefficients.

The final exercise involved dropping the state-specific trends. Generally, this had little effect on the signs of the age interactions; in some cases they became slightly larger in the no-trends specification. The insurance measures generally had more negative signs in this specification. This is consistent with the hypothesis that these state-specific linear trends are indeed a proxy for technology improvements in neonatal treatments over the period.

Overall, these various exercises suggest that the results for gestation, birth weight, and the 5-minute Apgar score are quite robust.

7 Conclusion

Reproductive technologies have improved radically since the introduction of the first fertility drugs in the late 1960s. These technologies make conception possible for many couples who otherwise would be unable to reproduce. Many of these technologies increase the probability of having a multiple birth, typically a more risky pregnancy. These technologies also provide women considering delaying reproduction with insurance against later infertility.

This study examines the effect of subsidies for use of advanced reproductive technologies on fertility outcomes (twin births), birth outcomes (for twins and singletons), and longer term child health (for twins). The subsidies are state mandates that health insurers cover or offer to cover infertility treatment. Living in a mandate state leads to about a 10 percent increase in the probability of twin births for mothers over 30 in the birth certificate data. The mandates lead to a slightly smaller increase in the probability of twin births in the PUMS data. Subsidized infertility treatment and increased use of this treatment by older women (as proxied by interactions between age indicators and the insurance mandates) are each associated with small, statistically significant, negative effects on birth outcomes such as gestation, birth weight, and the 5-minute Apgar score. However, preliminary findings looking at the longer-term impacts of being born in a mandate state for twins are mixed and rarely significant.

This paper contributes to the ongoing debate about birth selection in health economics. In previous work that examines birth selection, heterogeneity or selection among mothers and the impact of maternal investment on birth outcomes are expected to affect birth outcomes in the same direction. Here, the taxing nature of infertility treatments suggests that women who give birth due to use of ART may invest more in their pregnancies than do other women. Thus, these negative impacts of insurance mandates suggest that infertility treatment itself may be associated with adverse birth outcomes or that selection into childbearing of previously infertile women may lead to worse birth outcomes. These findings for birth outcomes suggest further investigation of longer term impacts of mandates on child health is warranted.

References

Abma, J., Chandra, A., Mosher, W., Peterson, L. & Piccinino, L. (1997), 'Fertility, family planning, and women's health: New data from the 1995 National Survey of Family Growth', *Vital Health Statistics* **23**(19).

ACOG Committee on Practice Bulletins—Obstetrics (2002), 'ACOG practice bulletin no. 38: Perinatal care at the threshold of viability', *Obstetrics and Gynecology* **100**(3), 618–23.

Almond, D., Chay, K. Y. & Lee, D. S. (2002), Does low birth weight matter? Evidence from the U.S. population of twin births, Working Paper 53, Center for Labor Economics, UC Berkeley.

Almond, D., Chay, K. Y. & Lee, D. S. (2004), The costs of low birth weight, Working Paper W10552, NBER.

Ananat, E. O., Gruber, J. & Levine, P. (2004), Abortion legalization and lifecycle fertility, Working Paper W10705, NBER.

Baker, L. C. & Phibbs, C. S. (2002), 'Managed care, technology adoption, and health care: The adoption of neonatal intensive care', *RAND Journal of Economics* **33**(3), 524–48.

Bao, Y. & Sturm, R. (2004), 'The effects of state mental health parity legislation on perceived quality of insurance coverage, perceived access to care, and use of mental health speciality care', *Health Services Research* **39**(5), 1361–263–275.

Becker, G. S. (1991), *A Treatise on the Family*, Harvard University Press, Cambridge, MA.

Behrman, J. R. & Rosenzweig, M. R. (2001), The returns to increasing body weight. Pier Working Paper 01-052.

Bergh, T., Ericson, A., Hillensjo, T., Nygren, K.-G. & Wennerholm, U.-B. (1999), 'Deliveries and children born after in-vitro fertilization in Sweden 1982–95: A retrospective cohort study', *Lancet* **354**(5), 1579–85.

Bertrand, M., Duflo, E. & Mullainathan, S. (2004), 'How much should we trust differences-in-differences estimates?', *Quarterly Journal of Economics* **119**(1), 249–275.

Buckles, K. (2004), Stopping the biological clock: Fertility therapies and the career/family tradeoff. Mimeo.

Callahan, T. L., Hall, J. E., Ettner, S. L., Christiansen, C. L., Greene, M. F. & Crowley, W. F. (1994), 'The economic impact of multiple-gestation pregnancies and the contribution of assisted-reproduction techniques to their incidence', *New England Journal of Medicine* **331**(4), 244–49.

Chandra, A. & Stephen, E. H. (1998), 'Impaired fecundity in the United States: 1982–1995', *Family Planning Perspectives* **30**(1), 34–42.

Cnattingius, S., Forman, M. R., Berendes, H. W. & Isotalo, L. (1992), 'Delayed childbearing and risk of adverse perinatal outcome: A population based study', *Journal of the American Medical Association* **268**(7), 886–90.

Colvin, M., McGuire, W. & Fowlie, P. (2004), 'Neurodevelopmental outcomes after preterm birth', *BMJ* **329**, 1390–1393.

Currie, J. & Hyson, R. (1998), Is the impact of health shocks cushioned by socio-economic status? The case of low birthweight. Mimeo, UCLA.

Cutler, D. M. & Meara, E. (2000), The technology of birth: Is it worth it?, *in* A. M. Garber, ed., 'Frontiers in Health Policy Research', Vol. 3, MIT Press, Cambridge, MA, pp. 33–67.

Dhont, M., Sutter, P. D., Martens, G. R. . G. & Bekaert, A. (1999), 'Perinatal outcome of pregnancies after assisted reproduction: A case-control study', *American Journal of Obstetrics and Gynecology* **181**(3), 688–695.

Division of Vital Statistics, NCHS (2000), 'Technical appendix from Vital Statistics of the United States 1998 Natality'.

Division of Vital Statistics, NCHS (Various years), 'Natality detail'.

Donohue, J. J. & Levitt, S. (2001), 'The impact of legalized abortion on crime', *Quarterly Journal of Economics* **116**(2), 379–420.

Donohue, J. J. & Levitt, S. (2004), 'Further evidence that legalized abortion lowered crime: A reply to joyce', *Journal of Human Resources* **39**(1), 29–49.

Griffin, M. & Panak, W. F. (1998), 'The economic cost of infertility-related services: An examination of the Massachussetts infertility insurance mandate', *Fertility and Sterility* **70**(1), 22–29.

Grossman, M. & Joyce, T. J. (1990), 'Unobservables, pregnancy resolutions, and birth weight production functions in New York City', *Journal of Political Economy* **98**(5, Part 1), 983–1007.

Gruber, J., Levine, P. & Staiger, D. (1999), 'Abortion legalization and child living circumstances: Who is the "marginal" child', *Quarterly Journal of Economics* **115**(1), 263–292.

Hamilton, B. H. & McManus, B. (2004*a*), Infertility treatment markets: The effects of competition and policy. Typescript.

Hamilton, B. H. & McManus, B. (2004*b*), Technology adoption and market structure: Evidence from infertility treatment markets. Typescript.

Jain, T., Harlow, B. L. & Hornstein, M. D. (2002), 'Insurance coverage and outcomes of in vitro fertilization', *New England Journal of Medicine* **347**(9), 661–666.

Jensen, G. & Morrisey, M. (1999), 'Employer-sponsored health insurance and mandated benefit laws', *The Milbank Quarterly* **77**(4), 425–459.

Joyce, T. (2004), 'Did legalized abortion lower crime?', *Journal of Human Resources* **39**(1), 1–28.

Kezdi, G. (2002), Robust standard error estimation in fixed-effects panel models. University of Michigan, Typescript.

Klick, J. & Markowitz, S. (2003), Are mental health insurance mandates effective?: Evidence from suicides, Working Paper W9994, NBER.

Levi, A. J., Raynault, M. F., Bergh, P. A., Drews, M. R., Miller, B. T. & Scott, R. T. (2001), 'Reproductive outcome in patients with diminished ovarian reserve', *Fertility and Sterility* **76**(4), 666–669.

Liu, Z., Dow, W. H. & Norton, E. C. (2004), 'Effect of drive-through delivery laws on postpartum length of stay and hospital charges', *Journal of Health Economics* **23**(1), 129–155.

Martin, J. A., Hamilton, B. E., Ventura, S. J., Menacker, F., & Park, M. M. (2002), 'Births: Final data for 2000', *National Vital Statistics Reports* **50**(5).

Martin, J. A. & Park, M. M. (1999), 'Trends in twin and triplet births: 1980–1997', *National Vital Statistics Reports* **47**(24).

McCarton, C. M., Wallace, I. F., Divon, M. & Vaughan, H. G. (1996), 'Cognitive and neurologic development of the premature, small for gestational age infant through age 6: Comparison by birth weight and gestational age', *Pediatrics* **98**(6), 1167–78.

McCormick, M. C. (1985), 'The contribution of low birth weight to infant mortality and childhood morbidity', *The New England Journal of Medicine* **312**(2), 82–90.

McCormick, M. C., Brooks-Gunn, J., Workman-Daniels, K., Turner, J. & Peckham, G. J. (1992), 'The health and developmental status of very low-birth-weight children at school age', *Journal of the American Medical Association* **267**(16), 2204–08.

Medical Research International, American Fertility Society Special Interest Group (1988), 'In vitro fertilization/embryo transfer in the United States: 1985 and 1986 results from the National IVF/ET Registry', *Fertility and Sterility* **49**(2).

National Center for Chronic Disease Prevention and Health Promotion, ASRM, SART, RESOLVE (2002), *1998 Assisted Reproductive Technology Success Rates National Summary and Fertility Clinic Reports*. Downloaded 10/30/2002.

National Conference of State Legislatures (2002), Insurance coverage for infertility therapy. Downloaded 10/11/2002.

Neumann, P. J., Gharib, S. D. & Weinstein, M. C. (1994), 'The cost of a sucessful delivery with in vitro fertilization', *The New England Journal of Medicine* **331**(4), 239–43.

Office of Technology Assessment (1988), *Infertility: Medical and Social Choices*, U.S. Government Printing Office, Washington, DC.

Pacula, R. & Sturm, R. (2000), 'Mental health parity: Much ado about nothing?', *Health Services Research* **35**(1 Part 2), 263–275.

Rees, J. M., Lederman, S. A. & Kiely, J. L. (1996), 'Birth weight associated with lowest neonatal mortality: Infants of adolescent and adult mothers', *Pediatrics* **98**(6), 1161–15.

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

Rosenzweig, M. R. & Schultz, T. P. (1983), 'Estimating a household production function: Heterogeneity, the demand for health inputs, and their effects on birth weight', *Journal of Political Economy* **91**(5), 723–746.

Royer, H. (2003), What all women (and some men) want to know: Does maternal age affect infant health? Typescript, University of California at Berkeley.

Saunders, K., Spensley, J., Munro, J. & Halasz, G. (1996), 'Growth and physical outcome of children conceived by in vitro fertilization', *Pediatrics* **97**(5), 688–92.

Schieve, L. A., Meikle, S. F., Ferre, C., Peterson, H. B., Jeng, G. & Wilcox, L. S. (2002), 'Low and very low birth weight in infants conceived with use of assisted reproductive technology', *New England Journal of Medicine* **346**(10), 731–737.

Schmidt, L. (2005), Effects of infertility insurance mandates on fertility. Typescript.

Sheiner, E., Shoham-Vardi, I., Hershkovitz, R., Katz, M. & Mazor, M. (2001), 'Infertility treatment is an independent risk factor for caesarean section among nulliparous women aged 40 and above', *Obstetrics and Gynecology* **185**(4), 888–891.

Society for Assisted Reproductive Technology, American Society for Reproductive Medicine (1999), 'Assisted reproductive technology in the United States: 1996 results generated from the American Society for Reproductive Medicine/Society for Assisted Reproductive Technology Registry', *Fertility and Sterility* **71**(5).

Society for Assisted Reproductive Technology, American Society for Reproductive Medicine (2002), 'Assisted reproductive technology in the United States: 1998 results generated from the American Society for Reproductive Medicine/Society for Assisted Reproductive Technology Registry', *Fertility and Sterility* **77**(1).

Summers, L. (1989), 'Some simple economics of mandated benefits', *American Economic Review Papers and Proceedings* **79**(2), 177–183.

Sutcliffe, A. G. (2004), 'The long-term paediatric outcomes of assisted reproductive technologies', *International Congress Series* **1266**, 359–366.

Tan, S.-L., Doyle, P., Campbell, S., Beral, V., Rizk, B., Brinsden, P., Mason, B. & Edwards, R. G. (1992), 'Obstetric outcome of in vitro fertilization pregnancies compared with normally conceived pregnancies', *American Journal of Obstetrics and Gynecology* **167**(3), 778–784.

Varma, T. R. & Patel, R. H. (1987), 'Outcome of pregnancy following investigation and treatment of infertility', *International Journal of Gynaecology and Obstetrics* **25**(2), 113–20.

Vohr, B., Wright, L., Dusick, A., Mele, L., Verter, J., Steichen, J., Simon, N., Wilson, D., Boryles, S., Bauer, C., Delaney-Black, V., Yolton, K., Fleisher, B., Papile, L.-A. & Kaplan, M. (2000), 'Neurodevelopmental and functional outcomes of extremely low birth weight infants in the National Institute for Child Health and Human Development Neonatal Research Network, 1993–1994', *Pediatrics* **105**(6), 1216–1226.

Wolpin, K. I. (1997), Determinants and consequences of the mortality and health of infants and children, *in* M. Rosenzweig & O. Stark, eds, 'Handbook of Population and Family Economics', Vol. 1 A, Elsevier Science, New York, NY, pp. 483–557.

Zhang, J., Meikle, S., Grainger, D. & Trumble, A. (2002), 'Multifetal pregnancy in older women and perinatal outcomes', *Fertility and Sterility* **78**(3), 562.

Table 1: Means of Birth Outcomes, Infant, and Mother Characteristics by Presence of Infertility Treatment Mandate, Sample of Twins, Natality Data

	No mandate for Infertility treatment	Mandate for Infertility treatment
Gestation (wks.)	35.91	35.74
Premat. (< 37 wks.)	0.48	0.51
Very premat. (< 32 wks.)	0.11	0.11
Birth weight (gms.)	2394	2395
Low BW (< 2500 gms.)	0.52	0.52
Very low BW (< 1500 gms.)	0.10	0.10
Extremely low BW (< 1000 gms.)	0.05	0.05
Five minute Apgar score (0–10)	8.49	8.57
Mother has 4 yr. degree	0.19	0.26
Mother black	0.19	0.16
Mother Hispanic	0.08	0.21
Mother Hispanic not reported or missing	0.19	0.02
Age of mother	27.40	28.68
HI for infert., state law	0.00	1.00
HI for infert., may incl. IVF , state law	0.00	0.47
HI for infert., excl. IVF, state law	0.00	0.53
N	1221629	462468

Summary statistics for twin births from Detailed Natality files for 1981–1999. Column 1 contains means for all twins during years when no infertility treatment mandate was in place during the year before birth and column 2 contains means for all twins during years when any infertility treatment mandate was in place during the year before birth. Statistics weighted to account for 50% sampling in some states before 1985.

Table 2: Mother's Age and Share of Twin Births to Women in Specific Age Groups by Presence of Infertility Treatment Mandate, Sample of Twins, Natality Data

	No mandate for Infertility treatment	Mandate for Infertility treatment
Age of mother	27.40	28.68
Mother aged $\leq$ 19	0.082	0.070
Mother aged 20–24	0.242	0.192
Mother aged 25–29	0.316	0.277
Mother aged 30–34	0.247	0.285
Mother aged 35–39	0.098	0.146
Mother aged 40–44	0.013	0.026
Mother aged 45 or older	0.001	0.004
N	1221629	462468

Mother's age and share of births born to women of various ages by time period for all twin births in sample from Detailed Natality files. Column 1 contains means for all twins during years when no infertility treatment mandate was in place during the year before birth and column 2 contains means for all twins during years when any infertility treatment mandate was in place during the year before birth. Statistics weighted to account for 50% sampling in some states before 1985.

Table 3: Means of Educational Attainment by Age and Presence of Infertility Mandate, Sample of Twins, Natality Data

	No mandate for Infertility treatment	Mandate for Infertility treatment
Mother ≤ 29		
Mother's educ. < 4 yr. degree	0.87	0.87
Mother has 4 yr. degree	0.13	0.13
N, mother's education reported	686828	24229
N, mother's education missing	94075	9044
Mother 30–34		
Mother's educ. < 4 yr. degree	0.63	0.60
Mother has 4 yr. degree	0.37	0.40
N, mother's education reported	269636	128061
N, mother's education missing	33406	3620
Mother 35–39		
Mother's educ. < 4 yr. degree	0.58	0.54
Mother has 4 yr. degree	0.42	0.46
N, mother's education reported	107738	65999
N, mother's education missing	12509	1671
Mother ≥ 40		
Mother's educ. < 4 yr. degree	0.54	0.43
Mother has 4 yr. degree	0.46	0.57
N, mother's education reported	17436	327
N, mother's education missing	1752	0.02

Mother's education and total number of twins for twin births by mother's age group and presence of state level infertility mandate from Detailed Natality files. Each panel presents means and Ns for a specific age group for educational attainment for all records reporting education as well as the number of records which did not report education. Column 1 contains means for all twins during years when no infertility treatment mandate was in place during the year before birth and column 2 contains means for all twins during years when any infertility treatment mandate was in place during the year before birth. Statistics weighted to account for 50% sampling in some states before 1985.

Table 4: States Enacting Laws Mandating Health Insurance Offer or Cover Infertility Treatment, 1981–2003

State	Year law	*IVF is*		*Mandate to insurers to*		*Law applies to*		
	Passed	Not excluded	Excluded	Cover	Offer	All firms	Non-HMOs	Only HMOS
Arkansas	1987	1	0	1	0	0	1	0
California	1989	0	1	0	1	1	0	0
Connecticut	1989	1	0	0	1	0	1	0
Hawaii	1987	1	0	1	0	1	0	0
Illinois	1991	1	0	1	0	1	0	0
Louisiania	2001	0	1	1	0	1	0	0
Maryland	1985	1	0	1	0	1	0	0
Massachusetts	1987	1	0	1	0	1	0	0
Montana	1987	1	0	1	0	0	0	1
New York	1990	0	1	1	0	0	1	0
New Jersey	2001	1	0	1	0	1	0	0
Ohio	1991	1	0	1	0	0	0	1
Rhode Island	1989	1	0	1	0	1	0	0
Texas	1987	1	0	0	1	1	0	0
West Virginia	1977	1	0	1	0	0	0	1

Shown are the date each state passed a law related to health insurance coverage of infertility (column 1), whether the coverage mandate includes IVF or does not specify (column 2), and whether the coverage mandate excludes IVF (column 3), whether the coverage is mandated (column 4), and whether the insurer is required to offer it to the customer (column 5), whether all firms are covered (column 6), whether HMOs are excluded from the law (column 7), and whether the law applies only to HMOs (column 8).

Table 5: Means, Twin Indicator and Mother's Age at Birth by Presence of Infertility Treatment Mandate, Sample of Twins and Singletons, PUMS

	No mandate for Infertility treatment	Mandate for Infertility treatment
Child is twin	0.031	0.034
Mother aged 20–24	0.239	0.208
Mother aged 25–29	0.321	0.287
Mother aged 30–34	0.246	0.271
Mother aged 35–39	0.098	0.130
Mother aged 40–44	0.018	0.028
Mother aged 45 or older	0.002	0.003
Mother high school grad., no college	0.282	0.244
Mother some college, no 4 yr. deg.	0.340	0.311
Mother has 4 yr. degree	0.239	0.249
N	2408743	853081

Summary statistics for singletons and twins from PUMS. Column 1 contains means for all singletons and twins in years with no mandate for infertility treatment the year before birth and and column 2 contains means for all singletons and twins in years with a mandate for infertility treatment the year before birth. Sample is all children aged 5–17 in combined 2000 5% and 1% PUMS whose mother is in the household, who are either the only child or one of two children their age, both of whom are a child or grandchild of the householder and share the same mother, and whose age and sex and relationship to the head are not allocated. Twins are identified as children of the same age, whose mother lives in the household, who share the same mother, and for whom the householder is their parent or grandparent. Statistics weighted.

Table 6: Means, Child Outcomes, and Mandate Variables, Sample of Twins, PUMS

	No mandate for Infertility treatment	Mandate for Infertility treatment
Long-lasting condition		
Blindness/deafness/vision or hearing impairment	0.012	0.010
Limited in basic physical activities	0.013	0.012
Physical, mental, or emot. cond. lasting $\geq$ 6 months causes		
Difficulty learning, remembering, or concentrating	0.049	0.038
Difficulty dressing, bathing, or getting around house	0.012	0.012
N	57396	14907

Summary statistics for twins from PUMS. Column 1 contains means for all twins in years with no mandate for infertility treatment the year before birth and and column 2 contains means for all twins in years with a mandate for infertility treatment the year before birth. Sample is all children aged 5–17 in combined 2000 5% and 1% PUMS whose mother is in the household, who are one of two children their age, both of whom are a child or grandchild of the householder and share the same mother, and whose age and sex and relationship to the head are not allocated. Twins are identified as children of the same age, whose mother lives in the household, who share the same mother, and for whom the householder is their parent or grandparent. Statistics weighted.

Table 7: Coefficients on Regressions of Determinants of Being a Twin, Interactions with Mother 30 or Older, Sample of Singletons and Twins, Natality Data

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30	Cover/offer and Cover/offer $*$ age ≥ 30
HI for infert.	-0.00183^{**}		
	(0.00082)		
HI for infert. $* \geq 30$	0.00302^{***}		
	(0.00096)		
HI for infert. may incl. IVF		-0.00190^{*}	
		(0.00101)	
HI for infert. may incl. IVF $* \geq 30$		0.00282^{**}	
		(0.00113)	
HI for infert. excl. IVF		-0.00175	
		(0.00130)	
HI for infert. excl. IVF $* \geq 30$		0.00319^{**}	
		(0.00139)	
HI must cover infert.			-0.00393^{***}
			(0.00110)
HI must cover infert. $* \geq 30$			0.00673^{***}
			(0.00109)
HI must offer infert. coverage			0.00029
			(0.00104)
HI must offer infert. $* \geq 30$			-0.00034
			(0.00108)

Coefficients on insurance mandate variables, regressions of determinants of birth being a twin. Sample is one in fifty random subsample of all singleton and twin births for 1981–1999. Each column represents one regression. The regression in column 1 has an indicator for any mandate that insurers covering/offering to cover infertility treatment and its interaction with being 30 or older, that in column 2 has indicators for mandates for insurers on infertility treatment that exclude IVF treatment or that may include IVF, and their interactions with being 30 or older, that in column 3 has indicators for mandates for insurers to cover or offer to cover infertility treatment, and their interactions with being 30 or older, Regressions include state of residence fixed-effects, year fixed-effects, month of birth fixed-effects, and state of residence linear time trends. Standard errors clustered at the state-by-year level. Regressions are weighted to account for the 50% sampling in some states before 1985. N is 1,439,525. Pre-reform mean of dependent variable is 0.0227, that for women 30 or older is 0.0295, and that for women 35 or older is 0.0308. $^{***}p < .01, ^{**}p < .05, ^{*}p < .10$.

Table 8: Coefficients on Regressions of Determinants of Being a Twin, Interactions with Mother 35 or Older, Sample of Singletons and Twins, Natality Data

Controls for	Any mandate and Any mandate $*$ age $\geq$ 35	IVF/no IVF and IVF/no IVF $*$ age $\geq$ 35	Cover/offer and Cover/offer $*$ age $\geq$ 35
HI for infert.	-0.00142*		
	(0.00080)		
HI for infert. $* \geq 35$	0.00580***		
	(0.00119)		
HI for infert. may incl. IVF		-0.00160	
		(0.00101)	
HI for infert. may incl. IVF $* \geq 35$		0.00672***	
		(0.00173)	
HI for infert. excl. IVF		-0.00123	
		(0.00124)	
HI for infert. excl. IVF $* \geq 35$		0.00509***	
		(0.00139)	
HI must cover infert.			-0.00247**
			(0.00110)
HI must cover infert. $* \geq 35$			0.00797***
			(0.00157)
HI must offer infert. coverage			-0.00016
			(0.00101)
HI must offer infert. $* \geq 35$			0.00381***
			(0.00144)

Coefficients on insurance mandate variables, regressions of determinants of birth being a twin. Sample is one in fifty random subsample of all singleton and twin births for 1981–1999. Each column represents one regression. The regression in column 1 has an indicator for any mandate that insurers covering/offering to cover infertility treatment and its interaction with being 35 or older, that in column 2 has indicators for mandates for insurers on infertility treatment that exclude IVF treatment or that may include IVF, and their interactions with being 35 or older, that in column 3 has indicators for mandates for insurers to cover or offer to cover infertility treatment, and their interactions with being 35 or older. Regressions include state of residence fixed-effects, year fixed-effects, month of birth fixed-effects, and state of residence linear time trends. Standard errors clustered at the state-by-year level. Regressions are weighted to account for the 50% sampling in some states before 1985. N is 1,439,525. Pre-reform mean of dependent variable is 0.0227, that for women 30 or older is 0.0295, and that for women 35 or older is 0.0308. $^{***}p < .01, ^{**}p < .05, ^{*}p < .10$.

Table 9: Selected Coefficients, Determinants of Being a Twin, Interactions with Mother 30 or Older, Sample of Singletons and Twins, PUMS

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30
HI for infert.	0.00155	
	(0.00127)	
HI for infert. $* \geq 30$	0.00146	
	(0.00122)	
HI for infert. may incl. IVF		0.00104
		(0.00188)
HI for infert. may incl. IVF $* \geq 30$		-0.00115
		(0.00163)
HI for infert. excl. IVF		0.00211
		(0.00167)
HI for infert. excl. IVF $* \geq 30$		0.00394**
		(0.00154)

Coefficients on insurance mandate variables, regressions of determinants of birth being a twin. Sample is all children aged 0–17 in combined 2000 5% and 1% PUMS whose mother is in the household, who are either the only child or one of two children their age, both of whom are a child or grandchild of the householder and share the same mother, and whose age and sex and relationship to the head are not allocated. Twins are identified as children of the same age, whose mother lives in the household, who share the same mother, and for whom the householder is their parent or grandparent. Each column represents one regression. The regression in column 1 has an indicator for any mandate that insurers covering/offering to cover infertility treatment and its interaction with being 30 or older, that in column 2 has indicators for mandates for insurers on infertility treatment that exclude IVF treatment or that may include IVF. and their interactions with being 30 or older. All regressions include state of birth fixed effects for the child, year fixed-effects, and state of birth linear time trends. Regressions also include indicators for the mother's education being high school, some college, a four-year degree; for the mother's age in 5 year groupings; for the child being black, Asian, American Indian, or other non-white; for the child being Hispanic; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted. N is 3,260,683. Pre-reform mean of dependent variable is 0.0310, that for women 30 or older is 0.0314, and that for women 35 or older is 0.0315. $^{***}p < .01,^{**}p < .05,^{*}p < .10$.

Table 10: Selected Coefficients, Regressions of Determinants of Gestation (Weeks), Sample of Twins, Natality Data

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30
HI mandate for infert.	-0.128***	
	(0.050)	
HI for infert. $* \geq 30$	-0.046**	
	(0.021)	
HI for infert., may incl. IVF		-0.089*
		(0.054)
HI for infert., may incl. IVF $* \geq 30$		-0.000
		(0.028)
HI for infert., excl. IVF		-0.172**
		(0.073)
HI for infert., excl. IVF $*$ mother ≥ 30		-0.085***
		(0.023)

Coefficients on indicators for a state's mandating health insurers cover/offer to cover infertility treatments and interactions of age indicators with those variables. Sample is twin births in states reporting outcome for 1981–99. Column 1 regression includes an indicator for any mandate on infertility treatment and its interaction with the mother being 30 or older, and column 2 regression includes an indicator for insurers to cover/offer infertility treatment including and excluding IVF and their interactions with an indicator for being ≥ 30. All regressions include state of residence fixed-effects, year fixed-effects, month of birth fixed-effects, and state of residence linear time trends. Regressions also include indicators for the mother having had 1, 2, 3, 4, 5 or more prior live births, or for this being unknown; for her education being high school, some college, a four-year degree, or being unknown/not reported; for the mother being black, Asian, American Indian, or other non-white; for the mother being Hispanic, or for that not being reported; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted to account for 50% sampling in some states before 1985. N is 1,636,418. $^{***}p < .01,^{**} p < .05,^{*} p < .10$.

Table 11: Selected Coefficients, Regressions of Determinants of Premature Birth (Gest. < 37 weeks), Sample of Twins, Natality Data

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30
HI mandate for infert.	-0.000	
	(0.004)	
HI for infert. $* \geq 30$	0.009***	
	(0.003)	
HI for infert., may incl. IVF		0.003
		(0.006)
HI for infert., may incl. IVF $* \geq 30$		0.005*
		(0.003)
HI for infert., excl. IVF		-0.003
		(0.006)
HI for infert., excl. IVF $*$ mother ≥ 30		0.012***
		(0.003)

Coefficients on indicators for a state's mandating health insurers cover/offer to cover infertility treatments and interactions of age indicators with those variables. Sample is twin births in states reporting outcome for 1981–99. Column 1 regression includes an indicator for any mandate on infertility treatment and its interaction with the mother being 30 or older, and column 2 regression includes an indicator for insurers to cover/offer infertility treatment including and excluding IVF and their interactions with an indicator for being ≥ 30. All regressions include state of residence fixed-effects, year fixed-effects, month of birth fixed-effects, and state of residence linear time trends. Regressions also include indicators for the mother having had 1, 2, 3, 4, 5 or more prior live births, or for this being unknown; for her education being high school, some college, a four-year degree, or being unknown/not reported; for the mother being black, Asian, American Indian, or other non-white; for the mother being Hispanic, or for that not being reported; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted to account for 50% sampling in some states before 1985. N is 1,636,418. $^{***}p < .01,^{**} p < .05,^{*} p < .10$.

Table 12: Selected Coefficients, Regressions of Determinants of Very Premature Birth (Gest. < 32 weeks), Sample of Twins, Natality Data

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30
HI mandate for infert.	0.010***	
	(0.003)	
HI for infert. $* \geq 30$	0.003**	
	(0.002)	
HI for infert., may incl. IVF		0.009**
		(0.004)
HI for infert., may incl. IVF $* \geq 30$		-0.000
		(0.002)
HI for infert., excl. IVF		0.011**
		(0.005)
HI for infert., excl. IVF $*$ mother ≥ 30		0.007***
		(0.002)

Coefficients on indicators for a state's mandating health insurers cover/offer to cover infertility treatments and interactions of age indicators with those variables. Sample is twin births in states reporting outcome for 1981–99. Column 1 regression includes an indicator for any mandate on infertility treatment and its interaction with the mother being 30 or older, and column 2 regression includes an indicator for insurers to cover/offer infertility treatment including and excluding IVF and their interactions with an indicator for being ≥ 30. All regressions include state of residence fixed-effects, year fixed-effects, month of birth fixed-effects, and state of residence linear time trends. Regressions also include indicators for the mother having had 1, 2, 3, 4, 5 or more prior live births, or for this being unknown; for her education being high school, some college, a four-year degree, or being unknown/not reported; for the mother being black, Asian, American Indian, or other non-white; for the mother being Hispanic, or for that not being reported; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted to account for 50% sampling in some states before 1985. N is 1,636,418. $^{***}p < .01,^{**} p < .05,^{*} p < .10$.

Table 13: Selected Coefficients, Regressions of Determinants of Birth Weight (grams), Sample of Twins, Natality Data

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30
HI mandate for infert.	-15.453*	
	(8.290)	
HI for infert. $* \geq 30$	-16.673***	
	(3.620)	
HI for infert., may incl. IVF		-5.903
		(7.927)
HI for infert., may incl. IVF $* \geq 30$		-7.849
		(4.996)
HI for infert., excl. IVF		-26.240**
		(12.609)
HI for infert., excl. IVF $*$ mother ≥ 30		-24.145***
		(3.926)

Coefficients on indicators for a state's mandating health insurers cover/offer to cover infertility treatments and interactions of age indicators with those variables. Sample is twin births in states reporting outcome for 1981–99. Column 1 regression includes an indicator for any mandate on infertility treatment and its interaction with the mother being 30 or older, and column 2 regression includes an indicator for insurers to cover/offer infertility treatment including and excluding IVF and their interactions with an indicator for being ≥ 30. All regressions include state of residence fixed-effects, year fixed-effects, month of birth fixed-effects, and state of residence linear time trends. Regressions also include indicators for the mother having had 1, 2, 3, 4, 5 or more prior live births, or for this being unknown; for her education being high school, some college, a four-year degree, or being unknown/not reported; for the mother being black, Asian, American Indian, or other non-white; for the mother being Hispanic, or for that not being reported; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted to account for 50% sampling in some states before 1985. N is 1,676,147. $^{***}p < .01,^{**}p < .05,^{*}p < .10$.

Table 14: Selected Coefficients, Regressions of Determinants of Low Birth Weight (< 2500 grams), Sample of Twins, Natality Data

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30
HI mandate for infert.	-0.000	
	(0.004)	
HI for infert. $* \geq 30$	0.010***	
	(0.002)	
HI for infert., may incl. IVF		-0.001
		(0.004)
HI for infert., may incl. IVF $* \geq 30$		0.005
		(0.003)
HI for infert., excl. IVF		0.001
		(0.006)
HI for infert., excl. IVF $*$ mother ≥ 30		0.014***
		(0.003)

Coefficients on indicators for a state's mandating health insurers cover/offer to cover infertility treatments and interactions of age indicators with those variables. Sample is twin births in states reporting outcome for 1981–99. Column 1 regression includes an indicator for any mandate on infertility treatment and its interaction with the mother being 30 or older, and column 2 regression includes an indicator for insurers to cover/offer infertility treatment including and excluding IVF and their interactions with an indicator for being ≥ 30. All regressions include state of residence fixed-effects, year fixed-effects, month of birth fixed-effects, and state of residence linear time trends. Regressions also include indicators for the mother having had 1, 2, 3, 4, 5 or more prior live births, or for this being unknown; for her education being high school, some college, a four-year degree, or being unknown/not reported; for the mother being black, Asian, American Indian, or other non-white; for the mother being Hispanic, or for that not being reported; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted to account for 50% sampling in some states before 1985. N is 1,676,147. $^{***}p < .01, ^{**}p < .05, ^{*}p < .10$.

Table 15: Selected Coefficients, Regressions of Determinants of Very Low Birth Weight (< 1500 grams), Sample of Twins, Natality Data

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30
HI mandate for infert.	0.011***	
	(0.003)	
HI for infert. $* \geq 30$	0.004**	
	(0.001)	
HI for infert., may incl. IVF		0.008**
		(0.004)
HI for infert., may incl. IVF $* \geq 30$		0.001
		(0.002)
HI for infert., excl. IVF		0.016***
		(0.005)
HI for infert., excl. IVF $*$ mother ≥ 30		0.006***
		(0.002)

Coefficients on indicators for a state's mandating health insurers cover/offer to cover infertility treatments and interactions of age indicators with those variables. Sample is twin births in states reporting outcome for 1981–99. Column 1 regression includes an indicator for any mandate on infertility treatment and its interaction with the mother being 30 or older, and column 2 regression includes an indicator for insurers to cover/offer infertility treatment including and excluding IVF and their interactions with an indicator for being ≥ 30. All regressions include state of residence fixed-effects, year fixed-effects, month of birth fixed-effects, and state of residence linear time trends. Regressions also include indicators for the mother having had 1, 2, 3, 4, 5 or more prior live births, or for this being unknown; for her education being high school, some college, a four-year degree, or being unknown/not reported; for the mother being black, Asian, American Indian, or other non-white; for the mother being Hispanic, or for that not being reported; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted to account for 50% sampling in some states before 1985. N is 1,676,147. $^{***}p < .01, ^{**}p < .05, ^{*}p < .10$.

Table 16: Selected Coefficients, Regressions of Determinants of Extremely Low Birth Weight (< 1000 grams), Sample of Twins, Natality Data

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30
HI mandate for infert.	0.010***	
	(0.003)	
HI for infert. $* \geq 30$	0.002	
	(0.001)	
HI for infert., may incl. IVF		0.005
		(0.003)
HI for infert., may incl. IVF $* \geq 30$		-0.000
		(0.001)
HI for infert., excl. IVF		0.016***
		(0.005)
HI for infert., excl. IVF $*$ mother ≥ 30		0.003***
		(0.001)

Coefficients on indicators for a state's mandating health insurers cover/offer to cover infertility treatments and interactions of age indicators with those variables. Sample is twin births in states reporting outcome for 1981–99. Column 1 regression includes an indicator for any mandate on infertility treatment and its interaction with the mother being 30 or older, and column 2 regression includes an indicator for insurers to cover/offer infertility treatment including and excluding IVF and their interactions with an indicator for being ≥ 30. All regressions include state of residence fixed-effects, year fixed-effects, month of birth fixed-effects, and state of residence linear time trends. Regressions also include indicators for the mother having had 1, 2, 3, 4, 5 or more prior live births, or for this being unknown; for her education being high school, some college, a four-year degree, or being unknown/not reported; for the mother being black, Asian, American Indian, or other non-white; for the mother being Hispanic, or for that not being reported; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted to account for 50% sampling in some states before 1985. N is 1,676,147. $^{***}p < .01,^{**} p < .05,^{*} p < .10$.

Table 17: Selected Coefficients, Regressions of Determinants of 5-Minute Apgar Score, Sample of Twins, Natality Data

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30
HI mandate for infert.	-0.038*	
	(0.020)	
HI for infert. $* \geq 30$	-0.029***	
	(0.008)	
HI for infert., may incl. IVF		-0.008
		(0.017)
HI for infert., may incl. IVF $* \geq 30$		-0.034***
		(0.010)
HI for infert., excl. IVF		-0.073**
		(0.034)
HI for infert., excl. IVF $*$ mother ≥ 30		-0.022**
		(0.010)

Coefficients on indicators for a state's mandating health insurers cover/offer to cover infertility treatments and interactions of age indicators with those variables. Sample is twin births in states reporting outcome for 1981–99. Column 1 regression includes an indicator for any mandate on infertility treatment and its interaction with the mother being 30 or older, and column 2 regression includes an indicator for insurers to cover/offer infertility treatment including and excluding IVF and their interactions with an indicator for being ≥ 30. All regressions include state of residence fixed-effects, year fixed-effects, month of birth fixed-effects, and state of residence linear time trends. Regressions also include indicators for the mother having had 1, 2, 3, 4, 5 or more prior live births, or for this being unknown; for her education being high school, some college, a four-year degree, or being unknown/not reported; for the mother being black, Asian, American Indian, or other non-white; for the mother being Hispanic, or for that not being reported; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted to account for 50% sampling in some states before 1985. N is 1,313,105. $^{***}p < .01,^{**} p < .05,^{*} p < .10$.

Table 18: Selected Coefficients, Regressions of Determinants of Gestation (Weeks), Sample of Singletons, Natality Data

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30
HI mandate for infert.	-0.020	
	(0.018)	
HI for infert. $* \geq 30$	-0.019*	
	(0.010)	
HI for infert., may incl. IVF		-0.004
		(0.023)
HI for infert., may incl. IVF $* \geq 30$		-0.031***
		(0.012)
HI for infert., excl. IVF		-0.037
		(0.023)
HI for infert., excl. IVF $*$ mother ≥ 30		-0.009
		(0.014)

Coefficients on indicators for a state's mandating health insurers cover/offer to cover infertility treatments and interactions of age indicators with those variables. Sample is 1 in 50 random sample of singleton births in states reporting outcome for 1981–99. Column 1 regression includes an indicator for any mandate on infertility treatment and its interaction with the mother being 30 or older, and column 2 regression includes an indicator for insurers to cover/offer infertility treatment including and excluding IVF and their interactions with an indicator for being ≥ 30. All regressions include state of residence fixed-effects, year fixed-effects, month of birth fixed-effects, and state of residence linear time trends. Regressions also include indicators for the mother having had 1, 2, 3, 4, 5 or more prior live births, or for this being unknown; for her education being high school, some college, a four-year degree, or being unknown/not reported; for the mother being black, Asian, American Indian, or other non-white; for the mother being Hispanic, or for that not being reported; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted to account for 50% sampling in some states before 1985. N is 1,366,019. $^{***}p < .01,^{**} p < .05,^{*} p < .10$.

Table 19: Selected Coefficients, Regressions of Determinants of Birth Weight (grams), Sample of Singletons, Natality Data

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30
HI mandate for infert.	-10.120**	
	(4.305)	
HI for infert. $* \geq 30$	-8.204***	
	(2.936)	
HI for infert., may incl. IVF		-7.171
		(5.531)
HI for infert., may incl. IVF $* \geq 30$		-1.488
		(3.580)
HI for infert., excl. IVF		-13.688**
		(5.705)
HI for infert., excl. IVF $*$ mother ≥ 30		-13.792***
		(3.556)

Coefficients on indicators for a state's mandating health insurers cover/offer to cover infertility treatments and interactions of age indicators with those variables. Sample is 1 in 50 random samples of singleton births in states reporting outcome for 1981–99. Column 1 regression includes an indicator for any mandate on infertility treatment and its interaction with the mother being 30 or older, and column 2 regression includes an indicator for insurers to cover/offer infertility treatment including and excluding IVF and their interactions with an indicator for being ≥ 30. All regressions include state of residence fixed-effects, year fixed-effects, month of birth fixed-effects, and state of residence linear time trends. Regressions also include indicators for the mother having had 1, 2, 3, 4, 5 or more prior live births, or for this being unknown; for her education being high school, some college, a four-year degree, or being unknown/not reported; for the mother being black, Asian, American Indian, or other non-white; for the mother being Hispanic, or for that not being reported; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted to account for 50% sampling in some states before 1985. N is 1,404,233. $^{***}p < .01,^{**} p < .05,^{*} p < .10$.

Table 20: Selected Coefficients, Regressions of Determinants of 5-Minute Apgar Score, Sample of Singletons, Natality Data

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30
HI mandate for infert.	-0.001	
	(0.009)	
HI for infert. $* \geq 30$	0.000	
	(0.006)	
HI for infert., may incl. IVF		0.019***
		(0.007)
HI for infert., may incl. IVF $* \geq 30$		-0.012**
		(0.006)
HI for infert., excl. IVF		-0.026*
		(0.014)
HI for infert., excl. IVF $*$ mother ≥ 30		0.014
		(0.010)

Coefficients on indicators for a state's mandating health insurers cover/offer to cover infertility treatments and interactions of age indicators with those variables. Sample is 1 in 50 random samples of singleton births in states reporting outcome for 1981–99. Column 1 regression includes an indicator for any mandate on infertility treatment and its interaction with the mother being 30 or older, and column 2 regression includes an indicator for insurers to cover/offer infertility treatment including and excluding IVF and their interactions with an indicator for being ≥ 30. All regressions include state of residence fixed-effects, year fixed-effects, month of birth fixed-effects, and state of residence linear time trends. Regressions also include indicators for the mother having had 1, 2, 3, 4, 5 or more prior live births, or for this being unknown; for her education being high school, some college, a four-year degree, or being unknown/not reported; for the mother being black, Asian, American Indian, or other non-white; for the mother being Hispanic, or for that not being reported; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted to account for 50% sampling in some states before 1985. N is 1,080,284. $^{***}p < .01, ^{**}p < .05, ^{*}p < .10$.

Table 21: Selected Coefficients, Regresssions of Determinants of Sensory Impairment, Sample of Twins, PUMS

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30
HI mandate for infert.	0.00222	
	(0.00340)	
HI for infert. $* \geq 30$	0.00281	
	(0.00288)	
HI for infert. may incl. IVF		0.00608
		(0.00502)
HI for infert. may incl. IVF. $* \geq 30$		0.00521
		(0.00393)
HI for infert. excl. IVF		-0.00169
		(0.00479)
HI for infert. excl. IVF $* \geq 30$		0.00082
		(0.00364)

Coefficients on insurance mandate variables, regressions of determinants of having long-lasting blindness, deafness, or a long-lasting severe vision or hearing impairment. Sample is twin children aged 5–17 in combined 2000 5% and 1% PUMS whose mother is in the household, who are one of two children their age, both of whom are a child or grandchild of the householder and share the same mother, and whose age and sex and relationship to the head are not allocated. Twins are identified as children of the same age, whose mother lives in the household, who share the same mother, and for whom the householder is their parent or grandparent. Each column represents one regression. The regression in column 1 has an indicator for any mandate that insurers covering/offering to cover infertility treatment and its interaction with the mother being 30 or older, that in column 2 has indicators for mandates for insurers on infertility treatment that exclude IVF treatment or that may include IVF and their interactions with the mother being 30 or older. All regressions include state of birth fixed effects for the child, year fixed-effects, and state of birth linear time trends. Regressions also include indicators for the mother's education being high school, some college, a four-year degree; for the mother's age in 5 year groupings; for the child being black, Asian, American Indian, or other non-white; for the child being Hispanic; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted. N is 72,303. Pre-reform mean of dependent variable is 0.0120, that for women 30 and older is 0.0119, and that for women 35 and older is 0.0117. $^{***}p < .01, ^{**}p < .05, ^{*}p < .10$.

Table 22: Selected Coefficients, Regresssions of Determinants of Physical Impairment, Sample of Twins, PUMS

Controls for	Any mandate and Any mandate ∗ age ≥ 30	IVF/no IVF and IVF/no IVF ∗ age ≥ 30
HI mandate for infert.	-0.00307	
	(0.00351)	
HI for infert. ∗ ≥ 30	0.00508	
	(0.00336)	
HI for infert. may incl. IVF		0.00104
		(0.00490)
HI for infert. may incl. IVF. ∗ ≥ 30		0.00408
		(0.00469)
HI for infert. excl. IVF		-0.00730
		(0.00500)
HI for infert. excl. IVF ∗ ≥ 30		0.00605
		(0.00430)

Coefficients on insurance mandate variables, regressions of determinants of having a long-lasting condition that severely limits one or more basic physical activities such as walking, climbing stairs, reaching, lifting, or carrying. Sample is twin children aged 5–17 in combined 2000 5% and 1% PUMS whose mother is in the household, who are one of two children their age, both of whom are a child or grandchild of the householder and share the same mother, and whose age and sex and relationship to the head are not allocated. Twins are identified as children of the same age, whose mother lives in the household, who share the same mother, and for whom the householder is their parent or grandparent. Each column represents one regression. The regression in column 1 has an indicator for any mandate that insurers covering/offering to cover infertility treatment and its interaction with the mother being 30 or older, that in column 2 has indicators for mandates for insurers on infertility treatment that exclude IVF treatment or that may include IVF and their interactions with the mother being 30 or older. All regressions include state of birth fixed effects for the child, year fixed-effects, and state of birth linear time trends. Regressions also include indicators for the mother's education being high school, some college, a four-year degree; for the mother's age in 5 year groupings; for the child being black, Asian, American Indian, or other non-white; for the child being Hispanic; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted. N is 72,303. Pre-reform mean of dependent variable is 0.0128, that for women 30 and older is 0.0123, and that for women 35 and older is 0.0114. $^{***}p < .01,^{**} p < .05,^{*} p < .10$.

Table 23: Selected Coefficients, Regresssions of Determinants of Condition Limiting Mental Activities, Sample of Twins, PUMS

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30
HI mandate for infert.	-0.00485	
	(0.01018)	
HI for infert. $* \geq 30$	0.00285	
	(0.00568)	
HI for infert. may incl. IVF		-0.01395
		(0.01574)
HI for infert. may incl. IVF. $* \geq 30$		0.00471
		(0.00872)
HI for infert. excl. IVF		0.00452
		(0.00973)
HI for infert. excl. IVF $* \geq 30$		0.00100
		(0.00627)

Coefficients on insurance mandate variables, regressions of determinants of having a physical, mental or emotional contion lasting at least 6 months that causes difficulty learning, remembering, or concentrating. Sample is twin children aged 5–17 in combined 2000 5% and 1% PUMS whose mother is in the household, who are one of two children their age, both of whom are a child or grandchild of the householder and share the same mother, and whose age and sex and relationship to the head are not allocated. Twins are identified as children of the same age, whose mother lives in the household, who share the same mother, and for whom the householder is their parent or grandparent. Each column represents one regression. The regression in column 1 has an indicator for any mandate that insurers covering/offering to cover infertility treatment and its interaction with the mother being 30 or older, that in column 2 has indicators for mandates for insurers on infertility treatment that exclude IVF treatment or that may include IVF and their interactions with the mother being 30 or older. All regressions include state of birth fixed effects for the child, year fixed-effects, and state of birth linear time trends. Regressions also include indicators for the mother's education being high school, some college, a four-year degree; for the mother's age in 5 year groupings; for the child being black, Asian, American Indian, or other non-white; for the child being Hispanic; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted. N is 72,303. Pre-reform mean of dependent variable is 0.0494, that for women 30 and older is 0.0479, and that for women 35 and older is 0.0478. $^{***}p < .01,^{**} p < .05,^{*} p < .10$.

Table 24: Selected Coefficients, Regresssions of Determinants of Condition Limiting Physical Activities, Sample of Twins, PUMS

Controls for	Any mandate and Any mandate $*$ age ≥ 30	IVF/no IVF and IVF/no IVF $*$ age ≥ 30
HI mandate for infert.	-0.00683	
	(0.00503)	
HI for infert. $* \geq 30$	0.00091	
	(0.00318)	
HI for infert. may incl. IVF		-0.00944
		(0.00786)
HI for infert. may incl. IVF. $* \geq 30$		-0.00301
		(0.00417)
HI for infert. excl. IVF		-0.00424
		(0.00459)
HI for infert. excl. IVF $* \geq 30$		0.00423
		(0.00405)

Coefficients on insurance mandate variables, regressions of determinants of having a physical, mental or emotional contion lasting at least 6 months that causes difficulty dressing, bathing, or getting around inside the house. Sample is twin children aged 5–17 in combined 2000 5% and 1% PUMS whose mother is in the household, who are one of two children their age, both of whom are a child or grandchild of the householder and share the same mother, and whose age and sex and relationship to the head are not allocated. Twins are identified as children of the same age, whose mother lives in the household, who share the same mother, and for whom the householder is their parent or grandparent. Each column represents one regression. The regression in column 1 has an indicator for any mandate that insurers covering/offering to cover infertility treatment and its interaction with the mother being 30 or older, that in column 2 has indicators for mandates for insurers on infertility treatment that exclude IVF treatment or that may include IVF and their interactions with the mother being 30 or older. All regressions include state of birth fixed effects for the child, year fixed-effects, and state of birth linear time trends. Regressions also include indicators for the mother's education being high school, some college, a four-year degree; for the mother's age in 5 year groupings; for the child being black, Asian, American Indian, or other non-white; for the child being Hispanic; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted. N is 72,303. Pre-reform mean of dependent variable is 0.0119, that for women 30 and older is 0.0116, and that for women 35 and older is 0.0114. $^{***}p < .01,^{**} p < .05,^{*} p < .10$.

Appendix Table 1: Means, State Controls, Sample of Twins, Natality Data

	All, 81–99	Twin matched, 81–99
% Hispanic	9.50	9.49
% black	12.51	12.53
Medicaid eligibility threshold as share of FPL	108.56	111.81
Real annual max. AFDC/TANF ben., family of 4 (97 $1000s)	6.46	6.40
Real median income, family of 4 (97 $1000s)	50.66	50.73
Overall unemployment rate (as share)	0.06	0.06
Employment growth rate (as share of employment)	0.02	0.02
Share under the poverty level	0.14	0.14
Share of births to unmarried women	0.28	0.28
Obstetric beds/1000 wom. 15–44	1.03	1.03
Neonatal intens. care beds/1000 wom. 15–44	0.19	0.19
Neonatal interm. care beds/1000 wom. 15–44	0.08	0.08
N	1684097	1456610

Summary statistics for all state level controls for sample of all twin births. Column 1 contains means for sample of all twins during 1981–99 and column 2 contains means for sample of all twins whose other twin was matched to them during 1981–99. Statistics weighted to account for 50% sampling in some states before 1985.

Appendix Table 2: Coefficients on Controls, Regressions of Determinants of Gestation (Weeks), Sample of Twins, Natality Data

Controls for	Any mandate and Any mandate $*$ age ≥ 35	IVF/no IVF and IVF/no IVF $*$ age ≥ 35
One previous live birth	0.366***	0.365***
	(0.008)	(0.008)
Two previous live births	0.703***	0.703***
	(0.011)	(0.011)
Three previous live births	0.732***	0.731***
	(0.014)	(0.014)
Four previous live births	0.672***	0.672***
	(0.019)	(0.019)
Five or more previous live births	0.681***	0.682***
	(0.026)	(0.026)
Birth order missing	-0.129	-0.127
	(0.125)	(0.125)
Mother high school grad., no college	0.102***	0.102***
	(0.016)	(0.016)
Mother some college, no 4 yr. deg.	0.166***	0.166***
	(0.017)	(0.017)
Mother has 4 yr. degree	0.243***	0.243***
	(0.018)	(0.018)
Mother's education missing	-0.448***	-0.457***
	(0.069)	(0.069)
Mother black	-0.970***	-0.970***
	(0.016)	(0.016)
Mother Asian or Pacific islander	0.139***	0.139***
	(0.027)	(0.027)
Mother American Indian	0.114**	0.113**
	(0.056)	(0.056)
Mother Hispanic	0.060***	0.060***
	(0.019)	(0.019)
No Hispanic indicator	-0.004	-0.000
	(0.040)	(0.040)
N		

Coefficients on other controls, regressions of determinants of gestation. Sample is all twin births for 1981–1999. Each column represents one regression. The regression in column 1 has an indicator for any mandate that insurers covering/offering to cover infertility treatment and its interaction with an indicator for the mother being ≥ 30, that in column 2 has indicators for mandates for insurers on infertility treatment that exclude IVF treatment or that may include IVF, and their interactions with an indicator for the mother being ≥ 30. Regressions include state of residence fixed-effects, year fixed-effects, month of birth fixed-effects, and state of residence linear time trends. clustered at the state-by-year level. Regressions are weighted to account for 50% sampling in some states before 1985. N is 1,636,418. $^{***}p < .01,^{**}p < .05,^{*}p < .10$.

Appendix Table 3: Selected Coefficients, Determinants of Being a Twin, Interactions with Mother 35 or Older, Sample of Singletons and Twins, PUMS

Controls for	Any mandate and Any mandate $*$ age ≥ 35	IVF/no IVF and IVF/no IVF $*$ age ≥ 35
HI for infert.	0.00164	
	(0.00121)	
HI for infert. $* \geq 35$	0.00377*	
	(0.00211)	
HI for infert. may incl. IVF		0.00052
		(0.00179)
HI for infert. may incl. IVF $* \geq 35$		0.00070
		(0.00238)
HI for infert. excl. IVF		0.00283*
		(0.00161)
HI for infert. excl. IVF $* \geq 35$		0.00640**
		(0.00306)

Coefficients on insurance mandate variables, regressions of determinants of birth being a twin. Sample is all children aged 0–17 in combined 2000 5% and 1% PUMS whose mother is in the household, who are either the only child or one of two children their age, both of whom are a child or grandchild of the householder and share the same mother, and whose age and sex and relationship to the head are not allocated. Twins are identified as children of the same age, whose mother lives in the household, who share the same mother, and for whom the householder is their parent or grandparent. Each column represents one regression. The regression in column 1 has an indicator for any mandate that insurers covering/offering to cover infertility treatment and its interaction with being 35 or older, that in column 2 has indicators for mandates for insurers on infertility treatment that exclude IVF treatment or that may include IVF. and their interactions with being 35 or older. All regressions include state of birth fixed effects for the child, year fixed-effects, and state of birth linear time trends. Regressions also include indicators for the mother's education being high school, some college, a four-year degree; for the mother's age in 5 year groupings; for the child being black, Asian, American Indian, or other non-white; for the child being Hispanic; for the child's gender; and for state-level economic, demographic, and public assistance variables. Standard errors clustered at the state-by-year level. Regressions are weighted. N is 3,260,683. Pre-reform mean of dependent variable is 0.0310, that for women 30 or older is 0.0314, and that for women 35 or older is 0.0315. $^{***}p < .01, ^{**}p < .05, ^{*}p < .10$.