The Impact of Extended Reproductive Time Horizons: Evidence from Israel's Expansion of Access to IVF

Naomi Gershoni* and Corinne Low†

December 1, 2016

Abstract

Women who delay childbearing to make time-costly career and educational investments face a lower probability of having a child, since women's fertility significantly declines with age. In addition, they may be penalized on the marriage market for their lower "reproductive capital" and end up with a lower-quality spouse. Israel's 1994 policy change to make *in vitro* fertilization and other assisted reproduction technologies free created an exogenous shock to later life fecundity, providing women with a form of insurance against age-related infertility. We use that policy change to study the impact of expected fertility decline on women's educational choices and marriage outcomes. We find that following the policy change, women are more likely to marry later, complete college education, and achieve post-college education. Moreover, after the change, the observed decrease in spousal quality for women who get married in their thirties rather than their twenties dissipates. This suggests that both men and women's decisions were affected by their updated perception of women's fertility prospects. More generally, our findings indicate that the asymmetry in later-life fertility between men and women is an important force in explaining women's educational, career, and marriage outcomes, and thus policies that protect against later life infertility can have far-reaching impacts.

1 Introduction

The introduction of technology that allowed women to delay fertility—"the pill"—has been tied to greater educational investments and improved labor market outcomes for women (Goldin and Katz, 2002; Bailey, 2006, 2010). However, women choosing to delay fertility in order to make career investments encounter a second biological constraint: a significantly lower probability of conception and successful birth.

*Tel Aviv University, The Eitan Berglas School of Economics, naomiger@post.tau.ac.il.

†University of Pennsylvania, The Wharton School, corlow@wharton.upenn.edu. We are grateful for the advice of Analia Schlosser, Pierre-Andre Chiappori, Cristian Pop-Eleches, and Yoram Weiss. We thank Tal Gross, David Weiss, Moshe Hazan, Alma Cohen, Neale Mahoney, Maya Rossin-Slater, and the participants at the NBER summer institute for helpful comments, and Jennie Huang and Ting Wang for excellent research assistance. We thank the staff of Tel Aviv University and the Israeli Central Bureau of Statistics for facilitating data access. The authors gratefully acknowledge research funding from the Wharton Dean's Research fund, the Wharton Pension Research Council/Boettner Center for Pensions and Retirement Research, and the Population Studies Center of the University of Pennsylvania. Opinions and views are solely the responsibility of the authors.

Women experience a sharp decline in fertility starting in their mid-thirties until a complete loss of fertility at menopause (unlike men, whose fertility declines gradually with age).[1] This gender-asymmetry in later life fertility is a possible source of inequality in outcomes and achievements between men and women, especially for careers with longer and more demanding initial investment.[2] Young women may be discouraged from making time-costly career and educational investments that delay marriage and childbearing. Moreover, women who make such investments may be penalized on the marriage market for their lower "reproductive capital," as shown by Low (2016). In light of the limited sources for exogenous shifts in women's later-life fertility, the role of this biological constraint in explaining women's career decisions and outcomes is not well understood.

Ideally, to study the impact of expected fertility on education decisions and marriage outcomes, we would randomly assign women to have longer or shorter periods of fecundity, and measure differences in educational investments, age at marriage, and quality of marriage match. In lieu of this, Israel's unprecedented decision to provide free *in vitro* fertilization (IVF) and other assisted reproductive technologies (ART) through the 1994 National Health Insurance Law provides a natural experiment: following the introduction of the new policy, women, and their prospective partners, could expect a longer period of fecundity.[3] In the ensuing years, Israeli families made wide use of these services. In 2002, for instance, 1,657 IVF treatment cycles per million people were performed in Israel, compared to 126 in the United States (Collins, 2002).[4]

We hypothesize that public access to IVF, which made it easier for older women to conceive, changed younger women's perceived cost of career investment. Consequently, they delayed marriage and pursued greater educational investments. Moreover, potential partners' estimation of women's fecundity horizons gradually changed, improving older women's marriage market outcomes. This was backed by a "thicker" marriage market for older women, as more women postponed marriage. Our theory of impact does not rely on the affected women actually using the technology themselves.

[1] Women lose 97% of eggs by 40 (Kelsey and Wallace, 2010), while remaining egg quality declines (Toner, 2003). The exact date of this decline may be difficult to pinpoint, but a collage of evidence points to pregnancies being rarer (Menken et al., 1986), more likely to end in miscarriage (Andersen et al., 2000), and more likely to result in fetal abnormalities (Hook et al., 1983) later in life, before the complete cessation of fecundity at some point far before men. The difficulty in measurement stems from the co-movement of fecundity and fertility choice, such as the use of contraceptives. Some literature uses couples in traditional societies that do not use birth control, but these measures may suffer from downward bias due to potentially declining rates of intercourse with age, and lower overall health and access to medical care in societies without contraceptive use. However, even more recent prospective studies show an accelerating decline in fecundity by age 40 for women, whereas men's fertility is relatively stable. For example, Rothman et al. (2013), in a prospective study of 2,820 Danish women trying to conceive, find that women 35-40 years old will become pregnant 77% as frequently as women age 20-24, whereas for men this ratio is 95%.

[2] One example of such "longer" investments would be attaining graduate education, but there are also on-the-job investments, like medical residencies, law firm partner tracks, and even the tenure track.

[3] In addition to the effect of the policy change itself, there was also a concurrent improvement in technology and widespread media attention to assisted reproduction surrounding the time of the policy change.

[4] We focus on IVF funding rather than other ARTs since this was the technology that most significantly affected chances of conception for older women. Moreover, costs of IVF are much higher than those of other infertility treatments, making funding a crucial determinant of usage. In the US for example, while costs of ovarian stimulatory drugs are a few hundred dollars, a single IVF treatment cycle costs around 10,000 dollars and the overall cost of IVF treatments per delivery is estimated to be higher than 50,000 dollars (Collins, 2001).

Rather, we propose that women, and their potential partners, view access to the technology as a form of insurance against age-related infertility. Even without specific knowledge of the policy change, the large amount of media attention to older women having children, and first-hand experience of observing motherhood at an older age, could have facilitated a *general* shift in beliefs regarding the time horizon of fertility.

We use data from the 2008 Israeli population census and a difference-in-differences strategy to examine the impact of the policy on women's decisions and outcomes. We examine the impact on women's marriage decisions, educational and career outcomes, and marriage market matching. First, we test whether women delayed marriage in response to their extended reproductive time horizons. We expect to find that women postpone marriage timing starting in 1994, when the new policy was announced, with the effect potentially increasing over time as awareness and willingness to rely on a relatively new medical technology builds up over time. To measure the impact, we compare women's age at first marriage to men's before and after 1994, presuming that men's marriage timing should not be impacted by the policy change (or at least impacted drastically less than women's via equilibrium effects). Prior to the policy change, men and women exhibit remarkably parallel trends for age at first marriage, making them a satisfactory comparison group.

We apply the same strategy to identify changes in women's rate of college and graduate school completion—comparing women's outcomes, which may be directly affected by women choosing to pursue additional education in the presence of IVF, to those of men, who should be affected much less (e.g., there could be a small general equilibrium effect if more educated women cause men to also increase their own education, but this should be much smaller than the direct effect). In this case, because we have individual-level data on educational attainment, we specify treatment by cohort: those entering either college or graduate school during the post-change period, rather than those already locked in to their educational choices. Despite the discontinuous nature of the change in IVF funding, we expect an evolving effect in this case because of the large variation in the age of university applicants and due to entry requirements which call for skill and effort (meaning not everyone will be marginal when they reach college-going age, and instead subsequent cohorts will benefit from the chance to adjust effort over time, prior to high school graduation). In addition, we use Arab-Israeli women, who were less likely then Jewish-Israeli women to use the technology and to be on the margin of large career investments, as an alternate control group, and find consistent effects.

Finally, we match data on women and their spouses to evaluate the change to "quality" of marriage matches for women who marry after age 30. Spousal quality is proxied by husband's income controlling for a variety of factors. Note that the gradual information dispersion may be especially salient for outcomes that involve equilibrium on the marriage market, which crucially relies on men's perceptions of potential wives' reproductive fitness and may take longer to update. We compare women who marry over thirty, and whose reproductive capital thus significantly increased due to the policy change, against women who marry younger and whose reproductive capital remained essentially unchanged.

We find statistically significant evidence that after the policy change women delayed marriage and completed more education. Women's age at first marriage increased immediately with the introduction of free IVF by 3 months on average, and kept increasing by approximately half a month on average per year for the next 14 years. Note that this effect could be driven by *some* women delaying marriage substantially in order to pursue greater educational investments, while others do not delay at all, an interpretation that is supported when we look at the distribution of marriage ages, and find it shifted those most like to marry after 22 but before 30. For college completion we find a gradual yet substantial increase of 0.7 percentage points per year, and approximately 2.5 percentage points difference between the pre and post periods. Graduate education increased 1.8 percentage points over a similar period of time, which is actually a larger proportional change than for college education, since a much smaller proportion of the population attends graduate school.[5] This last finding strongly supports our hypothesis, as later-life fertility is expected to have a larger impact for decisions that involve longer investments and hence longer delays. We additionally find increases in women's full-time employment and participation in prestigious occupations. Finally, we find a change in marriage outcomes for women who marry older, with increased spousal income for women who marry over age 30 compared to those who marry younger.

Our findings indicate that mitigating women's concerns for age-related infertility alters women's educational and marriage decisions. Furthermore, the marriage market responds to the change, in a way that measurably impacts matching along the dimension of spousal income, showing the dollars-and-cents value of reproductive capital. These results bolster the theory that the time-cost of education in terms of lost fertile years may be an important factor in women's educational decisions, and that fertility may be a valuable "asset" in attracting a more high-powered spouse on the marriage market. This research has implications for the economic understanding of women's career investment decisions, of the costs of aging to women, and of income inequality between genders. It can also inform an analysis of the welfare implications of a policy such as universal access to IVF, which appears to not only impact the women actually using IVF, but rather, affect all women by insuring against future infertility. Together, these findings point to the importance of biological differences in divergence of economic outcomes, and the potential role of policies in blunting this effect.

The remainder of the paper proceeds as follows: Section 2 discusses prior literature and the theoretical predictions; Section 3 describes the empirical setting for our project and the data we use; Section 4 presents results and tests their robustness, and Section 5 concludes.

[5] These estimates come from Table 2, columns 3 & 4, Table 4, columns 1-4, and Table 6, columns 1 & 2. We use the specifications that appear most suitable to describe the change in each outcome. To compare the estimated changes in college and graduate education completion, we use the standard difference-in-difference specification with group-specific time trends to account for a differential time trend evident in the pre-period.

2 Literature and Predictions

Previous literature establishes how important control over fertility is for women. Generally, this control can be divided into two types: for young women, control over fertility means being able to avoid unwanted pregnancies; however for older women, in their late thirties and forties, control over fertility means actually being able to conceive and give birth *when* they want to. While there is a vast body of empirical evidence on how "too much" fertility affects women's educational, career and marital prospects, the impact of "too little" fertility has not been sufficiently explored.

Goldin and Katz (2002) (and later Bailey (2006, 2010)) use the expansion of access to oral contraception to demonstrate that the ability to delay motherhood enabled women to make greater educational and labor market investments. Numerous additional studies support these findings, and use various methods to establish and quantify the tradeoff between family and career for women (Loughran and Zissimopoulos, 2009; Buckles, 2008; Blackburn et al., 1993; Taniguchi, 1999; Gustafsson, 2003; Miller, 2011; Avellar and Smock, 2003; Wilde et al., 2010). Additional recent work connects raising children to substantial wage declines for women (Adda et al., 2011; Kleven et al., 2015).

On the other hand, women who choose to delay fertility in favor of career investments risk achieving their desired family size. In addition to consequences for their own utility, this could lead to problems on the marriage market, since fecundity may be a trait that potential spouses value (Siow, 1998; Dessy and Djebbari, 2010; Bronson and Mazzocco, 2015). Low (2016) uses an online experiment to demonstrate that when age is randomly assigned to dating profiles, men, but not women, prefer younger partners. Interestingly, this preference is only true for men who have no children of their own and have accurate knowledge of the age-fertility relationship for women. An accompanying theoretical model shows that this can cause women to make lower career investments than they otherwise would have, and create worse marriage matches for women who do invest. Data on marriages in the United States during the 70s and 80s, supports this prediction, with women who obtained graduate education marrying poorer spouses than their college-educated peers.

This literature suggests that technology alleviating the problem of age-related infertility could increase women's educational investment and improve marriage outcomes for women who do invest. However, the research on assisted reproduction technology has, to date, mainly focused on outcomes of women who actually use the technology, rather than younger women who perceive it as offering insurance for infertility later in life. A series of papers uses the variation in the mandated insurance coverage of assisted reproductive technology (ART) across US states and over time to determine how more coverage affects IVF usage and outcomes (Velez et al., 2014; Hamilton and McManus, 2012; Bitler and Schmidt, 2012, 2006; Bundorf, Henne, and Baker, Bundorf et al.; Buckles, 2013; Schmidt, 2007, 2005), offering suggestive evidence that when coverage goes up, more women use IVF, fertility rates for older mothers go up, and multiple births rise.

A much more limited literature explores the impact of such mandates on the timing of marriage and childbearing, supporting the hypothesis that infertility treatments that will primarily be helpful later in life may influence the decisions of younger women (again using state-year variation). Ohinata (2011) finds that infertility insurance mandates resulted in 1-2 year delays in first birth among highly educated white women, and Abramowitz (2012, 2014) shows that increased access is associated with marriage delays for white women. The only evidence on career outcomes comes from Buckles (2007), which finds suggestive evidence that infertility insurance mandates led to increased labor force participation for women.

The approach of using state-year variation in IVF coverage mandates has limitations, especially when discussing general equilibrium shifts in perceptions of both men and women. Since these are small and localized policy changes, awareness may not be widespread, particularly with young women who may not even be managing their own insurance yet. More importantly, there is mixed evidence on how state health insurance mandates influence the insurance and labor market equilibrium: mandates may increase insurance premiums more significantly for the most affected workers and therefore negatively affect their wages and employment (Lahey (2012)).[6] Nonetheless, these papers find effects even with this more limited variation, and thus suggest an important potential contribution in testing the hypothesis that access to IVF may affect women's early-life decisions using a more discrete policy event.

The Israeli policy change thus provides a unique opportunity, in applying equally to all, and being widely discussed publicly. Moreover, given that the coverage is publicly funded, there are no concerns that the observed changes in women's career investment are driven by a shift in employers' costs and preferences for employing older women.

Thus, our paper offers the first opportunity to study a large-scale policy change that changed not just the actual chance of getting pregnant when older, but, crucially, the beliefs about this chance by both young women considering career investments and men considering marrying them later on. Moreover, this is the first paper to empirically study the impacts of a shift in later-life fertility potential on outcomes resulting from the decision to delay childbearing, including educational investments and the marriage match quality.

[6]With any employer-provided insurance benefit, wages may fall to reflect the presence of the benefit to employees, and cost to employers. Lahey (2012) presents evidence on infertility mandates suggesting that in addition to these falling wages, because wage changes will not fully offset the increased premium costs for women in affected age groups, employment opportunities (and thus labor force participation) for this group decreases.

3 Setting and Empirical Approach

3.1 IVF in Israel

Since the emergence of IVF technology in the early 80s, Israel has been on the forefront of IVF research: the first Israeli "test tube baby," born in 1982, was only the fifth IVF birth worldwide. However, until the early nineties, usage of the technology was still relatively low, and technological advances were slow in coming. IVF treatments were covered at least to some extent by four health plans, which provided medical insurance to 95% of the Israeli population,[7] but the extent of coverage and terms of eligibility varied between health plans and, in many cases, were vague or a priori undetermined. [8]

Following a widely covered, public debate, the Knesset enacted the 1994 National Health Insurance Law (NHI), which included IVF tests and treatments in a "basket" of free health services that all health plans must provide. The law provides all Israeli citizens with guaranteed access to:

> IVF treatments for the purpose of the birth of two children for couples who do not have children from their present marriage, as well as for childless women who wish to establish a single-parent family.[9]

The law, as originally written, did not place any restrictions on the age of women, or the number of attempts that could be made, and provided coverage for up to two "take-home babies". This is in stark contrast to most IVF coverage policies, which usually entitles beneficiaries to a certain number of *treatments*, rather than a certain outcome. The 1994 law thus provided access to IVF that is unmatched anywhere in the world, ushering in an era of expanded usage and technological improvement.

We argue that the 1994 law can be used as a quasi experiment, providing an exogenous shock to women's expected later life fertility. Importantly, the passing of the law was driven by the well-recognized "overtly pronatalist" Israeli agenda, rooted in the Jewish tradition of familism, rather

[7]The four health plans were partially subsidized by the government, but mostly relied on the membership fees they collected. Approximately 5% of the population, especially young people, had no health insurance at all.

[8]The most generous coverage was offered by the largest health plan ("Clalit"), which placed almost no limitations on usage. Interestingly, this generosity was a result of "faulty computer infrastructure (that could not trace women's treatment and entitlement efficiently) rather than from professional or social conviction" (Birenbaum-Carmeli, 2004). The other health plans offered a limited number of treatment cycles and placed age restrictions and long qualification periods. For example in the "Leumit" health plan the number of treatment cycles was limited to six and the maximal age was 40 (an exceptions committee was authorized to approve up to 3 cycles for women aged 40 to 42, two cycles for women aged 42-44, and one cycle for women aged 45). In all health plans, egg donations required approval by an exceptions committee. There was practically no competition among the different funds since transferability was highly limited, and membership was often the result of family legacy and political agenda—thus, customers were unlikely to "shop" for the best coverage.

[9]In practice, public funding covers approximately 85% of total treatment costs. Private and complementary health insurance programs of the health plans offer additional coverage, and also cover treatments for third and fourth child.

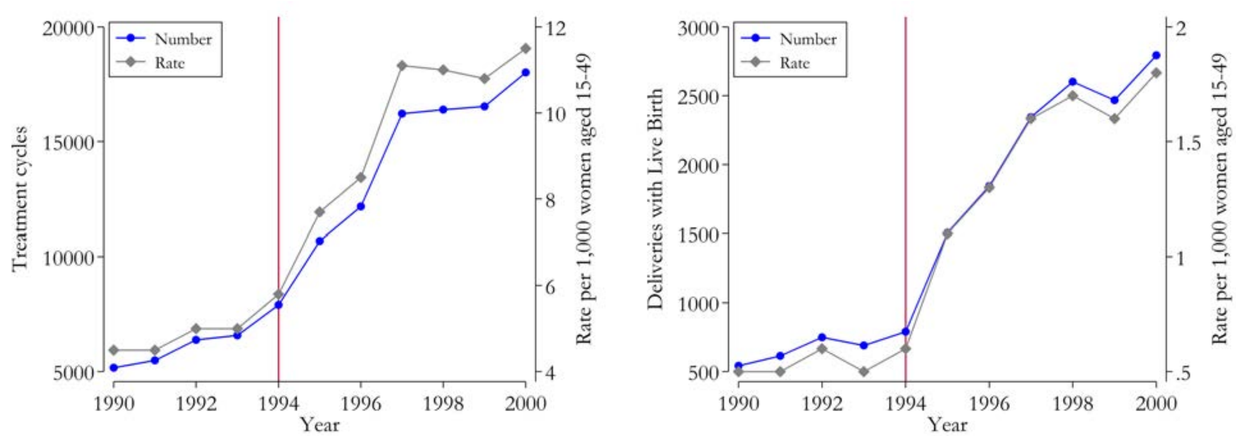
than pro-woman or "feminist" impulses which may have carried other effects.[10] As an example of the policy objectives behind the law, Israel's supreme court ruled (during numerous debates over the implications of the policy) that becoming a parent is a fundamental human right.[11]

The new and unique Israeli funding policy facilitated fast adoption and increased usage of the new fertility-enhancing technologies. Figure 1 shows that the number of IVF treatment cycles more than doubled in the 6 years following the approval of the new policy.[12] Although the benefits of the law came into effect in 1995, the increase in the number of IVF treatment cycles began already in 1994, with the large amount of press coverage and increased knowledge on IVF availability.[13] The figure on the right hand side shows that in the year after usage increased, there was a sharp increase in live deliveries using IVF.

Figure 1: Direct Impacts of IVF Access

(a) IVF Treatment Cycles

(b) IVF Deliveries with Live Births



Notes: Administrative data from Israeli Ministry of Health, covering all women in Israel.

Figure 2 shows the direct impact on older women's fertility, by measuring the increase in women over 40 (a significant portion of whom would require fertility technology of some kind to be able to conceive) with children under one year old. The data for this figure is from the Israeli Annual Labor Force Survey ("LFS data"), for the years 1980 to 2011. This graph shows a large jump in older motherhood in 1995, the year after women started using the technology, as shown in Figure

[10] Other examples of such policies are governmental child allowances and maternity grants, broad legal protection of working mothers' rights, extended funding of prenatal care and various tax benefits for parents. For an elaborate discussion of those policies and their evolvement over time, see Birenbaum-Carmeli (2003).

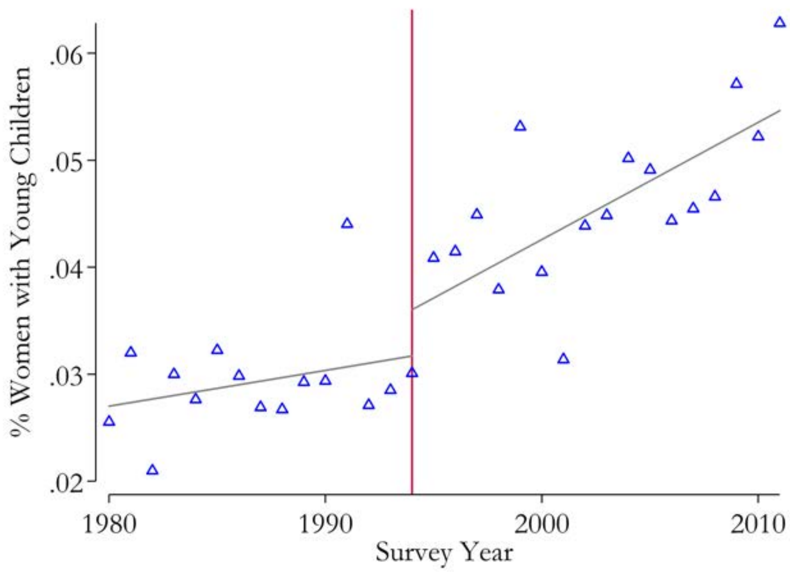
[11] See for example High Court 7052/03 Adalla vs. Ministry of Interior.

[12] The Israeli parliament "Kneset" issued a report in 2012 that attributes this dramatic change to the regularization and expansion of IVF funding under the NHI law.

[13] The common measure of usage is the number of IVF treatment cycles relative to the size of fertile women population. Since there is no documentation of the number of women treated each year, it is impossible to assess whether the sharp increase in usage stems from an increase in the number of women undergoing IVF treatments, or from an increase in the number of attempts each IVF patient makes. It is reasonable to assume that it is a result of a combination of these two, especially given the large increase in IVF-assisted births.

1.

Figure 2: Percentage of Women over 40 with Children ≤1 year, Labor Force Survey



Notes: The figure presents the percentage of women above age 40 (>40 and <=47, since in practice very few women above 47 have young children) with children of age 1 or below. Data from the Labor Force Survey 2001-2011, restricted to Israeli-born Jews.

In the years following the policy change, additional rules and regulations were enacted, standardizing practices surrounding IVF and its funding, and therefore supporting ongoing expansion of IVF usage.[14] Nowadays, there are 26 IVF clinics spread throughout Israel. Most public hospitals have an IVF unit, making treatment very easily accessible for most residents of Israel. Israel is the world leader in the rate of IVF treatment cycles and in the percentage of babies born following IVF treatments: approximately 4% of all babies born in Israel are conceived using IVF (Hashiloni-Dolev, 2013).[15]

During this time, the media was flooded with IVF success stories, such as extreme cases of women having children at advanced ages, further raising awareness of the new technology.[16] In Israeli press, local success stories were celebrated as “national accomplishments and symbols of local scientific excellence” (Birenbaum-Carmeli, 2004). The IVF law was itself heavily covered in

[14]The most distinct example is the 1996 Embryonic Carrying Agreement Law, officially legalizing and regulating surrogacy for the first time in the world (Simonstein, 2010).

[15]Compared to approximately 1-2% of the children born in other countries where IVF use is prevalent. The annual number of IVF cycles per million persons in Israel is the highest in the world and amounts to almost 3,500, compared to 2,000 in Denmark which is second (Birenbaum-Carmeli, 2010)

[16]For example, “World record: woman aged 60 gave birth to girl, Yedioth Aharonoth 22.2.94”; After 44 failed test-tube fertilizations, a 60-year-old woman gave birth to a baby girl in 1994.

the press, and continued to be covered as debates ensued on whether to limit coverage.[17]

The three forces of improved access, technology improvement, and publicity reinforced each other, leading to an IVF boom in Israel. This course of events drove a rapid and ongoing change in Israelis' attitudes and perceptions regarding IVF success rates, and thus the fertility time horizons for women. In fact, studies show that the widespread knowledge of IVF through media coverage (as well as personal experience with relatives or friends having successful older-age births may have even led to an over-estimate of later life fertility success rates (Hashiloni-Dolev et al., 2011).[18]

This technological and legal shift would have been particularly salient to Israelis, who place a strong emphasis on marriage and family relative to other countries. Israelis tend to marry young and have large families relative to other OECD countries.[19] Moreover, as Israelis tend to complete education later than in other OECD countries, due to mandatory military service, even early educational and career investments, such as completing college, may infringe on a woman's planned reproductive years, and potentially limit family size.[20] This makes Israel an ideal setting for the study of the impact of extended reproductive time horizons offered by IVF technology on women's decisions and outcomes.

3.2 Empirical Approach and Data

The 1994 Israeli IVF policy change provides a unique advantage over previously studied IVF policies, because it offers an opportunity to study the impact of an exogenous, generalized shift in beliefs about fertility horizons (rather than local and likely small changes from insurance mandates in the United States). The disadvantage inherent in our setting is that all Israelis received the update to beliefs, i.e. "treatment", at the same time, and thus there is no spatial variation that can be used for identification. We believe this tradeoff is warranted, as some of the most important impacts of extending reproductive time horizons may only occur when a policy affects widespread perceptions, thus affecting young people's choices, and future partners' perceptions on the marriage market. In

[17]The Ministry of Health expressed its intent to limit coverage to seven treatment cycles and provoked public protest. The press covered this conflict using personal stories of women over 40 that had children only following dozens of IVF treatment cycles and others who are still trying after a number of failures (Birenbaum-Carmeli, 2004).

[18]Hashiloni-Dolev et al. (2011) examines Israeli students' knowledge regarding age-dependent fertility decline, and finds a significant overestimation of the likelihood of pregnancy, especially for women over 40, with or without the aid of IVF.The study compared the students' estimations, measured by survey in 2009, to medical data. Unfortunately, there is no similar data collected prior to 1994. Similar studies in other countries also found overestimation of conception probability but not necessarily for the over 40 age group. In addition, those studies did not specifically target IVF success rates (see for example Bretherick et al. (2010)). In a different study, one third of the students that participated marked as "correct" statements declaring that healthy women over 45 have good chances of naturally conceiving and that the birth of a first child could be delayed till a woman turns 43 (Haimov-Kochman et al., 2012).

[19]For example, according to UN data, between 2000-2005 Israel's total fertility rate was 2.91, the highest among all 35 OECD current member countries, and far above the OECD average of 1.65. According to OECD data, in 2009 the marriage rate in Israel was 6.3 per 1000 residents compared to an OECD average of 5.0. Also, according to the Israeli Central Bureau of Statistics report "Women and men in Israel 1990-2009," the median age of first marriage in Israel is 3-4 years lower than in other western countries.

[20]According to OECD data, in 2011 Israel's median age at first graduation was slightly above 27 whereas the OECD average was slightly below 25.

the absence of spatial variation, we employ a difference-in-differences strategy, comparing groups which are expected to be more versus less affected by the policy, within the country.

We examine the impact of IVF on three main outcomes: age at marriage, higher education attainment, both college and graduate, and spousal quality for older women. Because of the obligatory military service, Israeli women end up entering college on average 4-5 years later than women in the United States. That, combined with younger marriage and higher total fertility rates in Israel (discussed in the previous section), means even college-going decisions may be affected by fertility constraints.

Our data comes from the 2008 Israeli population census, covering approximately 20% of Israeli households.[21] We present most of the analysis that follows for Israeli-born Jews, given that other population groups may have responded differently to the changes in IVF policy (Remennick, 2010). (Islam does not support all ART practices (especially surrogacy and ova donations) and the Roman Catholic church bans all types of ART(Birenbaum-Carmeli, 2003).)[22] Later, we utilize the differential response by Arab-Israelis to the IVF policy change to further explore the robustness of our results. We begin our analysis with an inspection of age at first marriage by year of marriage, using men as the "less affected" comparison group, over a 30-year study period, from 1979 to 2008. As aforementioned, men do not experience the same drop in fertility with age as women and therefore IVF funding does not affect their expectations for age-related levels of fecundity. This empirical strategy relies on men's and women's age at first marriage exhibiting parallel trends pre-intervention, which we demonstrate. Although men's marriage timing could be affected in equilibrium by shifts in women's choices, these effects would not exceed the initial impact on women. This does mean, however, that our effect should be considered a lower bound on the true impact. We consider 1994 to be the first year of the treatment period, as our treatment is knowledge of IVF availability and the resulting change of expectations, rather than the actual funding change.[23]

We first show a standard difference-in-differences specification, that measures the average difference in age at first marriage (AFM) between the "pre" and "post" periods, according to the following equation:

$$AFM = \beta_0 + \beta_1 fem + \beta_2 post + \beta_3 fem \times post + \beta_4 time + X'\gamma + u$$

[21]The survey began at the end of 2008 and was concluded in July 2009.

[22]In addition, there are significant differences between Jewish and Arab Israelis (which are the majority of the non-Jewish population) and between native born and immigrants in norms surrounding family and gender issues (see for example Danziger and Neuman (1999)). It is important to note that intermarriages between Jews and other religions are extremely rare in Israel. Moreover, in our sample and relevant period, the percentage of marriages between native-born Israelis and immigrants is very low. This is partially due to a large incoming flow of immigrants from the Former Soviet Union during the nineties which allowed this population to form segregated communities, with institutions designed to retain distinct cultural characteristics of this community.

[23]Our "treatment" kicks in as soon as women became aware of the insurance coverage rather than when coverage was actually implemented. Due to the large amount of press surrounding the approval of the law in 1994, we find it to be the turning point in the perceptual change regarding female later life fertility, although the law only came into effect in the beginning of 1995.

where X is a vector of individual level controls, which includes indicators for religiosity (ultra-orthodox or not) and parents' origin (Europe, Asia & Africa or Israeli born), to account for demographic shifts over time. We then test for both a change in levels at the time of the policy change and a change in the time-trend of the outcome variable, allowing us to examine the evolution of the effect over time:

$$\begin{aligned} AFM = \beta_0 + \beta_1 fem + \beta_2 post + \beta_3 fem \times post + \beta_4 time \\ + \beta_5 post \times time + \beta_6 fem \times time + \beta_7 fem \times post \times time + X'\gamma + u \end{aligned}$$

We also estimate these two equations with year of marriage fixed effects, to account for transitory shocks that may affect the marriage market.

Similarly, for college and graduate education, we compare women's outcomes to those of men.[24] However, for educational outcomes, we go by cohort and consider as treated the cohorts that were still at the relevant age for educational decisions at the time they learned about the increased access to IVF. As our main specification, we use the median age of applicants at the relevant year (as reported in macro data) to indicate the first treated cohort. We start with the 1951 birth cohort, but need to restrict our sample of younger birth cohorts to avoid censoring in educational outcomes among individuals who may not have completed their education by the time of the 2008 Census.[25] We use 1977 as our end-date for college education, therefore analyzing individuals no younger than 31 years old, and 1974 as the end date for graduate education, therefore analyzing individuals no younger than 34.

Since there is a slight upward trend in women's college education relative to men's pre-treatment, we control for a gender-specific linear time trends in the standard D-i-D specification, and estimate the following equation:

$$Education = \beta_0 + \beta_1 fem + \beta_2 post + \beta_3 fem \times post + \beta_4 time + \beta_5 fem \times time + X'\gamma + u$$

In addition, we estimate a specification which allows for changes both in level and trend (same as for age at first marriage) and add year of birth fixed effects to both specifications.

Because we may be concerned about other changes that could affect women's outcomes distinctly from men's, or general equilibrium effects, for these outcomes we employ a second strategy. We use female Arab-Israelis as a second control group, allowing us to account for any policies that may have affected all Israeli women differently than men. Arab-Israeli women make a suitable control group first and foremost because they are much less likely to use IVF, due to stronger religious

[24] It should be noted that just as with age at first marriage, men's education levels may be affected in general equilibrium (e.g., if partner education is complementary in marital surplus, and so more educated women leads to more educated men), but should be affected much less than women.

[25] We choose 1951 since it is the first cohort to have a reasonable number of observations for Israeli-born Jews (The state of Israel was founded in 1948).

restrictions on its use. In addition, the average Arab woman tends to marry younger and have a higher total fertility rate, compared to the average Jewish woman, which makes her less likely to be on the margin for time-costly career investments. [26]

Additionally, Arab women begin college on average 3-4 years younger than Jewish women since they do not serve in the military, creating a lower concern for the impact of college education on fertility.[27] Obviously, this alleviates their concerns for future fecundity. Another advantage of using this alternative comparison group is that, unlike Jewish men, there are unlikely to be general equilibrium effects on Arab women, because they Arab and Jewish Israelis have largely separate marriage markets. While this strategy may have its own potential confounding factors, they should be orthogonal to any issues presented by the male control group. Thus, if we estimate similar effects using the two strategies, it is unlikely that they are both caused by an omitted factor, rather than a reflection of the true treatment effect.

Finally, for spousal quality, our identification relies on the differential effects of additional fertile years for women who marry between 25 and 29 versus women who marry between 30 and 34. We place the cutoff between the two groups at 30 to exploit the perceived discontinuity in expected fertility exhibited at this age. Following the IVF policy change, the group of "older" brides (and their potential spouses) may expect greater fecundity following the policy change, while the marital fecundity of the "younger" brides remains essentially unchanged. We measure the impact of this difference over year of marriage, expecting to observe a lag in this outcome's response to the policy change, since reaching a new equilibrium in the marriage market takes time. Nevertheless we use 1994, the year of the new policy introduction, as the first year in the post-treatment period.

The main measure we use for spousal quality is income. Since our data is cross-sectional, we control for age effects by measuring spousal-income as the residual of a regression of income on a flexible polynomial in age.[28] On top of that, we restrict our sample period to 20 years between 1984-2003, in order to avoid measuring income of students or retirees. To confirm that our comparison is valid and insensitive to this choice, we further restrict our sample and repeat the estimation for the period between 1988-1999 (6 years pre and post "treatment"). Due to this restriction and to the expected lag in the showing of the new marriage market equilibrium, we have too few post-treatment periods to identify a change in trend (which for this outcome is very moderate to begin with) and thus focus on a difference-in-differences analysis for levels only:

[26] According to annual data published by the Isreli central bureau of statistics, at 1993 (just before the policy change) the median age at first marriage for Arab-Muslim Israeli women was 20 compared to 23.3 for Jewish Israeli women (average age was 21.1 compared to 24). Total fertility rates for Arab-Muslim women in Israel during the 90s were stable, at 4.5-4.7 births per women, compared to slightly below 3 for Jewish-Israeli women (as reported in CBS working paper no. 60, Fertility among Jewish and Muslim Women in Israel, by Ahmad Hleihel).

[27] See for example CBS report "Arabs in Higher Education in Israel - First Year Students for First Degree in 2011/12" issued October 21st 2014. It should also be noted that the variance of the age of college applicants is much larger for the Jewish population (based on CBS data processed and presented by Mr. Aviel Kranzler, Higher Education and Science department at the CBS.

[28] Education as an alternative measure of spousal quality is presented in the appendix.

$$Spouse\_inc = \beta_0 + \beta_1 older + \beta_2 post + \beta_3 older \times post + \beta_4 time + X'\gamma + u$$

Note, that the controls we use here are the same as before and address characteristics of the bride (although the outcome measure is for the groom). In the appendix we present results for specifications with additional controls.

For this outcome, we can also use men as a second control group in a triple-differences specification:

$$\begin{aligned} Spouse\_inc = {} & \beta_0 + \beta_1 older + \beta_2 post + \beta_3 older \times post \\ & + \beta_4 fem + \beta_5 older \times fem + \beta_6 post \times fem + \beta_7 older \times post \times fem + X'\gamma + u \end{aligned}$$

The coefficient of interest here is β_7 which can be interpreted as the change in spousal quality for "older" brides relative to younger brides, compared to those same differences in spousal quality for *grooms* marrying over this time period. This helps determine that the evolution we observe is driven by a female specific change, and not by some general shift for older marriages.

For all of the outcomes and specifications described above, we use two methods to calculate standard errors and present both. First, we cluster at the year $\times$ group level, to account for cross-sectionally correlated outcomes within each time period. Our second set of standard errors is clustered at the geography $\times$ group level, to account for potential serial correlation. Since the chief alternative explanations for our results have to do with shocks to women's access to education, which occur locally (e.g., schools being built), we believe this geographical clustering will be well equipped to deal with other potential sources of unrelated error correlation in women's outcomes (examples of similar sub-group clustering can be found in Agarwal et al (2015), clustering at the product level, and Hanlon (2015), clustering at the patent level). We use the standard definition of "natural regions" in Israel, assuming that our outcomes should have a random disturbance component which is dependent on this association of the individual.[29] These standard errors will help limit the tendency of difference-in-differences analyses to over-reject the null hypothesis due to underlying serial correlation in the data (Bertrand et al., 2004).

To further address the concern for serial correlation, we then re-estimate our main specification for each outcome using Generalized Least Squares, allowing for correlation both across and within time periods (as in Chandra, Gruber, McKnight, 2010), and collapsing our data into group-year cells. Finally, in the appendix, we show permutation tests for each of our main results, demonstrating that our effects are "large" relative to the actual variation present in the data.

[29]Natural regions were first defined by the Israeli central bureau of statistics in 1961 and updated over time. In the relevant period for the census data we use, there were 51 natural region, not including the occupied territories. Natural regions are "areas which are (as much as possible) continuous, unified and homogenous both in terms of physical structure, climate and soil and in terms of demographic, economic and social characteristics of the population.

Because there may have been other long-term societal trends that could have divergent effects for men and women, we perform several analyses to provide further evidence that the 1994 IVF policy change drives our results. First, we use a Quandt Likelihood Ratio (QLR) test[30] to search over all possible break dates, and show that our "treatment year" is indeed identified as the break among candidate dates. Second, we use event study graphs, charting the impact over time around the time of the policy change, to show that a pre-trend is not driving our results, but rather that the observed effects only become significant after the policy change. We additionally present a second control strategy exploiting the lower expected usage of IVF by Arab-Israeli women, as discussed above, and a variety of placebo tests. Finally, we rely on the specific combination of outcome variables to bolster the evidence that IVF access is the driver behind the changes. While there are a few other mechanisms that may have an effect on one of the outcomes we study, none of those can be expected to impact *both* women's educational and marriage decisions *and* marriage outcomes for older women. We review candidate alternative explanations in more detail in section 4.3.

Table 1 shows summary statistics for our sample, comparing Jewish women to Jewish men (our main comparison) as well as Arab women, and then women who marry while young (25-29) to those who marry older (30-34). In addition to showing means for our key outcomes and controls, Table 1 also compares pre-trends in outcomes for the different groups. Our specifications control for group-specific trends when needed.

4 Results

4.1 Women's Decisions

We first examine the impact of the policy on women's decisions regarding marriage timing and education. As noted above, we do this using a difference-in-differences framework, where women's outcomes before and after the policy change are compared to men's outcomes before and after.

Age at First Marriage The first outcome we examine is women's marriage timing. If women indeed feel more confident about their reproductive prospects later in life, they may be more willing to delay marriage, which could in turn allow greater rates of college and graduate education. More control over future fertility prospects may allow women to delay marriage by alleviating not just their own concerns for being able to have kids when they want to, but also for their marital prospects after having delayed.

In contrast to the other outcomes we consider, the decision to delay marriage is completely controlled by the individual and does not require meeting certain standards or going through some process, such as being accepted to a university (and, unlike changes to spousal quality, it does not

[30] See Andrews (1993).

Table 1: Summary Statistics

Marrying pre-1994:	Jewish women N=38,370		Jewish men N=33,949		Arab women N=14,901	
	Mean	SD	Mean	SD	Mean	SD
Ultra-Orthodox	0.09	0.28	0.09	0.29	N/A	N/A
European-born mother	0.24	0.43	0.28	0.45	N/A	N/A
Asian/African-born mother	0.54	0.50	0.52	0.50	N/A	N/A
Income (Shekels)	95,629	92,393	186,756	173,543	53,299	49,999
Age	44.81	5.39	47.46	5.34	42.24	5.66
Age at first marriage	23.15	3.91	25.86	3.86	21.03	4.11
AFM pre-trend (SE)	0.13	(0.00)	0.14	(0.00)	0.08	(0.01)
College age pre-1994:	N=61,000		N=58,704		N=22,278	
	Mean	SD	Mean	SD	Mean	SD
College Educated	0.32	0.47	0.29	0.45	0.09	0.28
College pre-trend (SE)	0.0057	(0.00)	0.0041	(0.00)	0.0035	(0.00)
Grad-school age pre-1994:	N=46,428		N=44,355		N=16,449	
	Mean	SD	Mean	SD	Mean	SD
Highly Educated	0.12	0.32	0.11	0.31	0.01	0.11
Highly Ed. pre-trend (SE)	0.0009	(0.00)	0.0008	(0.00)	0.0033	(0.00)

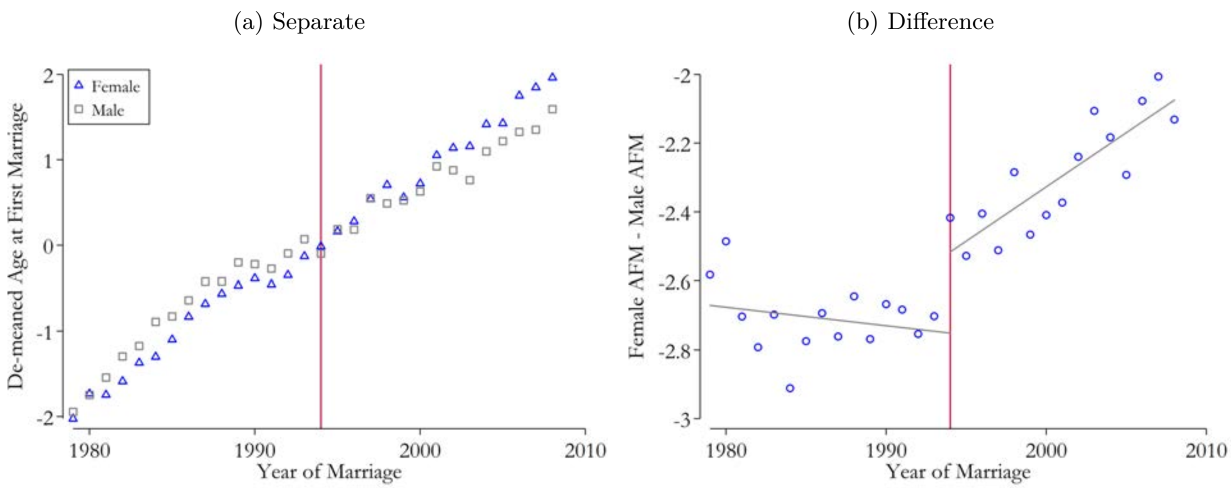
Marrying pre-1994:	Married 30-34 N=3,549		Married 25-29 N=11,227	
	Mean	SD	Mean	SD
Ultra-Orthodox	0.03	0.18	0.03	0.17
European-born mother	0.33	0.47	0.25	0.43
Asia/African-born mother	0.24	0.43	0.29	0.46
Income (Shekels)	92,238	91,189	99,701	96,225
Spousal Income	165,845	169,070	192,490	182,346
Sp. Income pre-trend (SE)	168	(1,277)	461	(697)

Notes: 2008 Israeli population census (20% sample). Restricted to Israeli-born. Sample "marrying pre-1994" is those married 1979 - 1993, inclusive. "College age pre-1993" is those born 1951 - 1970 for Jewish population, and 1954 - 1973 for Arab population. "Grad-school age pre-1994" is those born 1951 - 1966 for Jewish population, 1954 - 1969 for Arab population. Table by marriage age uses sample of women married 1984 - 1993 (shorter range due to income censoring for older individuals) with spousal matches in 2008 census.

require shifts in men's beliefs as well).[31] This should enable us to identify a clean and immediate effect precisely at the year when the policy was introduced.

Here, time represents the year the marriages are taking place, and thus, the treatment year is 1994, the year of the policy change. Figure 3 clearly shows that pre-trends for men and women were parallel, implying that women's marriage age was practically constant relative to men's until 1994. Note that men and women's age at first marriage also appears to respond to common shocks, in the lefthand figure that graphs outcomes separately (with age at first marriage de-meaned). Starting in 1994, women's marriage age increases relative to men. The graph of the difference in outcomes shows a sharp increase in level immediately at 1994, followed by a substantial positive change in trend.

Figure 3: Female vs. Male Age at First Marriage



Notes: Figure (a) shows average age at first marriage for women and men, by year of marriage, de-meaned so that the relative changes can be seen more clearly. Figure (b) presents the difference in average age at first marriage between women and men, as well as fitted lines for the pre (1979-1993) and post (1994-2008) periods. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

Table 2 analyzes this change using a regression, in both a simple difference-in-differences framework (columns 1 and 2), and an analysis demonstrating the change both in level and trend (columns 3 and 4). The latter indicates an increasing change in the outcome over time, which correlates with the gradual change in perceptions, rather than a one-time jump. We find that women marry about a third of a year older on average, relative to men, after the policy change. Both the discontinuity as well as the slope change are significant, although at a lower level under the geographically clustered standard errors. Columns 5 and 6 show the General Least Squares specification, where data is collapsed to year-group cells, for the D-i-D with slopes. Once again, both the intercept and

[31] It should also be noted that in Israel, couples tend to marry very soon after becoming engaged, and so there is not an extensive "lag" between the decision to get married and marriage itself.

slope change at 1994 are significant.

Table 2: Age at First Marriage

	Dependent variable: Age at First Marriage					
	DiD		Slope-Change DiD		GLS Slope-Change DiD	
	(1)	(2)	(3)	(4)	(5)	(6)
fem × post	0.412	0.415	0.241	0.246	0.247	0.195
	(0.073)***	(0.036)***	(0.127)*	(0.044)***	(0.082)***	(0.058)***
	[0.221]*	[0.220]*	[0.135]*	[0.133]*		
fem × post × time			0.039	0.040	0.035	0.036
			(0.013)***	(0.005)***	(0.010)***	(0.007)***
			[0.021]*	[0.022]*		
fem × time			-0.008	-0.008	-0.009	-0.007
			(0.012)	(0.003)**	(0.008)	(0.005)
			[0.015]	[0.015]		
post × time			0.001	0.013	-0.002	0.042
			(0.010)	(0.008)	(0.013)	(0.004)***
			[0.014]	[0.020]		
post	-0.440		-0.322		-0.395	
	(0.084)***		(0.091)***		(0.108)***	
	[0.141]***		[0.079]***			
female	-2.649	-2.651	-2.710	-2.715	-2.783	-2.757
	(0.059)***	(0.016)***	(0.114)***	(0.031)***	(0.067)***	(0.045)***
	[0.236]***	[0.235]***	[0.253]***	[0.250]***		
time	0.176	0.165	0.168	0.158	0.138	0.110
	(0.004)***	(0.004)***	(0.008)***	(0.004)***	(0.010)***	(0.002)***
	[0.011]***	[0.011]***	[0.011]***	[0.009]***		
Constant	26.319	26.030	26.251	25.896	26.856	26.275
	(0.070)***	(0.074)***	(0.096)***	(0.071)***	(0.088)***	(0.021)***
	[0.198]***	[0.221]***	[0.188]***	[0.164]***		
FEs		YES		YES		YES
Observations	167416	167416	167416	167416	60	60
R-Squared	0.246	0.246	0.246	0.247		

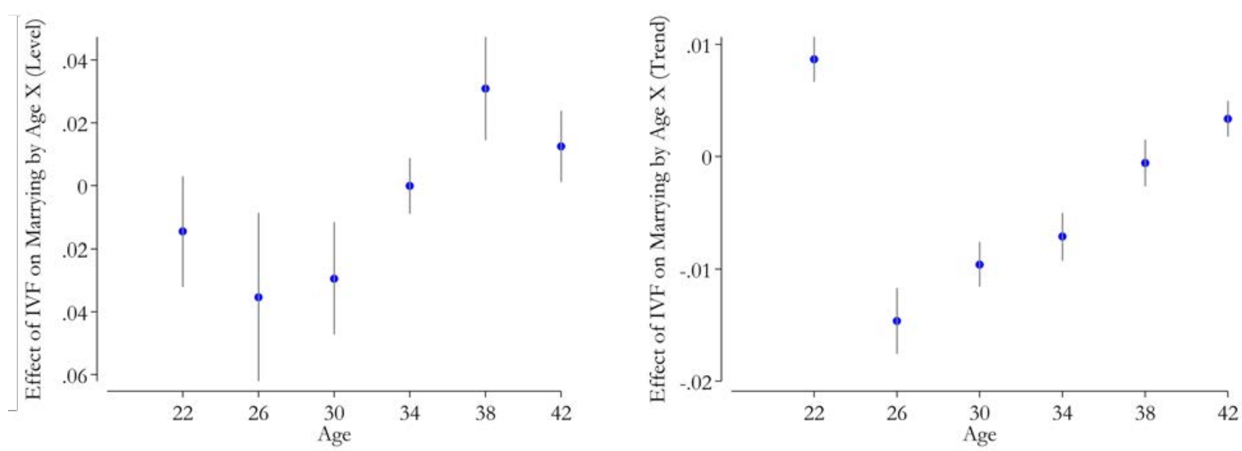
Notes: Columns 1–4: Ordinary least-squares difference-in-differences regression using micro data, including controls for religiosity and parents' origin. Robust standard errors clustered at the gender × year level in parentheses; robust standard errors clustered at the gender × geography level in square brackets. Columns 5–6: Generalized least squares regression with data collapsed to the gender-year of marriage level. Robust standard errors that allow for cross-sectional correlation and for serial correlation with panel specific correlation parameter, in parentheses. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

To understand how the distribution of marriage age was affected, rather than just the average, we run a series of regressions using the column 3 specification, but replacing the outcome variable with an indicator for being married by a certain age. Figure 4 shows the point estimates and confidence intervals for the two coefficients of interest on the interaction terms $fem{\times}post$ and $fem{\times}post{\times}time$, for each age cutoff. Figure 4(a) presents estimates for the immediate change in level (i.e. change in the percentage of women married by the specified age) and figure 4(b) shows the estimated change in trend. The two graphs show no decrease in marrying by age 22, which provides a useful falsification test, since we would not expect women inclined to marry and begin childbearing by age 22 to be concerned about fertility in their late thirties, and hence to be affected by access to IVF. We see the largest reduction in marriage by age 26, and from there a steadily decreasing impact, until the total effect reaches zero at age 38. The lack of reduction in marriage by age 38 suggests that women are delaying marriage, but not forgoing it entirely. Overall, this analysis suggests that the decrease in average marriage age after the policy change is mostly driven by women delaying marriages from their mid- and late-twenties into their thirties and even late thirties.

Figure 4: Married by Age Coefficients

(a) Intercept coefficient

(b) Slope coefficient



Notes: The figures presents the point estimates and confidence intervals of the coefficients on (a) the interaction term fem$\times$post and (b) the interaction term fem$\times$post$\times$time, for regressions where the outcome is a binary variable indicating whether or not the individual got married at or before a certain age, and the specification is as in column (3) in table 2. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

Quandt Likelihood Ratio breakpoint test To confirm that what we are picking up is truly a discontinuous shift in age at first marriage—a break in the time series—rather than more gradual time trends, we perform a Quandt Likelihood Test to "search" for the most likely break year in the data, over our entire sample period except for 15% "trimming" on either end, to account for limited data at the beginning and end of the sample period. We perform this test for age at first marriage, because it is the outcome measure that should have the cleanest "break" at 1994, since the education outcomes rely on cohorts entering in 1994, which may be imprecise, and spousal income relies on shifts in the marriage market, which may take more time.

To implement the test, we run a loop of regressions identical in specification to our columns 3 and 4 regressions, except the "break" year changes in each regression. We then perform an F-test for whether the two "break" parameters—slope and intercept—are different from zero. Finally, we search for the maximal F-stat among these tests.

As shown in Table 3, the test returns the highest F-statistic for 1994, which indicates that the year of the policy change and hence our treatment year is the most probable break year. The procedure for the QLR specifies comparing this "sup F-stat" to a table of critical values adjusted for the number of tests–the critical value for two restrictions and 15% trimming is 5.86, whereas the QLR statistic for age at first marriage for the "break" year is 10.38 or 10.78, depending on whether fixed effects are used or not.

Table 3: Quandt Likelihood Ratio test for break point

	F-Statistic	
Year of Marriage	No FEs	With YoM FEs
1983	7.05	7.40
1984	7.28	7.67
1985	7.29	7.69
1986	7.40	7.85
1987	7.83	8.27
1988	8.06	8.35
1989	8.12	8.36
1990	8.09	8.27
1991	8.03	8.16
1992	7.83	8.11
1993	8.57	8.91
1994	**10.38**	**10.78**
1995	7.19	7.49
1996	6.49	6.75
1997	5.61	5.86
1998	5.97	6.14
1999	3.95	4.07
2000	4.52	4.69
2001	4.02	4.16
2002	3.95	4.03
2003	2.60	2.64
2004	1.03	1.04

Notes: Table reports F stats from a regression according to the specification in columns 3 and 4 of Table 2, where the hypothesis is that the coefficients on *post* × *fem* and *post* × *fem* × *time* equal 0, and "post" is defined as being greater than or equal to the indicated year. Standard errors are not clustered in this case, as clustering is not conventional in QLR models, but similar results are obtained with clustering.

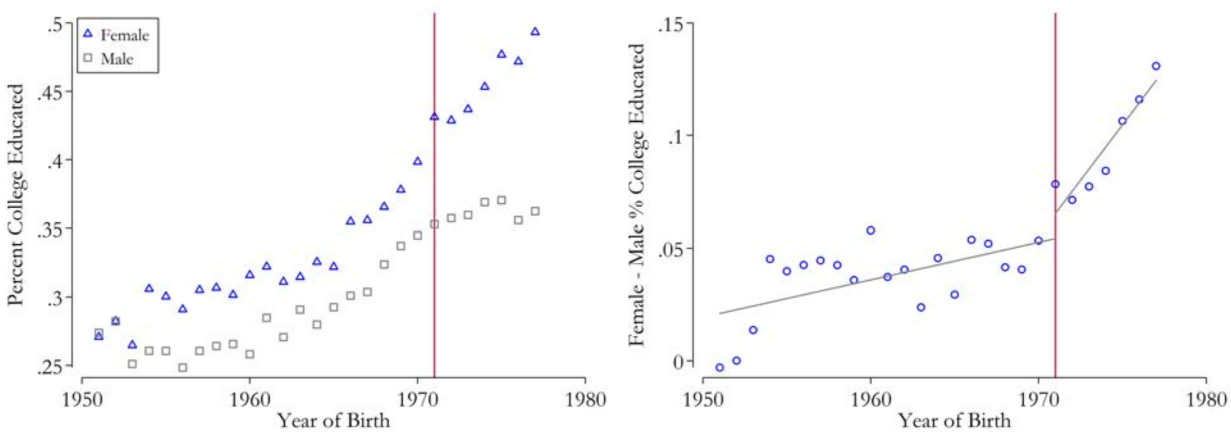
College Education We then turn to women's educational investments. Figure 5 shows the raw data used in this analysis, charting women's college completion compared to men, by year of birth.

The median age for college entry for women in Israel is 22.5.[32] The significant difference versus US patterns of college entry are due to both two years of mandatory military service (which may start "off cycle," and thus cause some individuals to delay longer) in addition to entry exams that are typically taken in a transition year between the military and college entry. Thus, we use the cohort born in 1971 as the first treatment year, as they would have been 23 and thus still able to be influenced in completing college by the change to their reproductive time horizon. Most women born in earlier cohorts would have already been past the age of making decisions about college completion at the time of the policy change.[33]

Figure 5: Difference in Percentage of College Educated Female and Male

(a) Separate

(b) Difference



Notes: Figure (a) shows average college attainment for women and men by birth cohort. Figure (b) presents the difference in college attainment between women and men, as well as fitted lines for the pre (1951-1970 birth cohorts) and post (1971-1977 birth cohorts) periods. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

Figure 5 shows that men and women's college education moves roughly in parallel prior to the 1994 time change seemingly responding to common shocks, with the exception of the earliest three cohorts, in which a different pattern is present, creating a slight upward pre-trend in women's education relative to men's. We therefore include group-specific time trends in the difference-in-differences regression specification. Following the policy change, there is a shift upward in women's college attainment, which continues as an upward slope change. Note that because we cannot pinpoint the 1971 cohort as the "treated" cohort, as we can with those who marry in 1994, it is certainly possible that some individuals from earlier cohorts were still able to get additional college education in response to the policy change, and thus in the appendix we present figure 5 with an interval around the treatment period (see figure 19).

[32]as reported by the Israel Central Bureau of Statistics, and confirmed in 1995 Census data for the cohort of interest

[33]Men's median age for college entry is 24. Because men enter college slightly later, we experiment with shifting the treatment year for men as one of our robustness checks, shown in the appendix. See table A1.

These results are presented formally in a regression in Table 4. Columns 1 and 2 show estimates for a simple difference-in-differences specification, with men as the control group and gender-specific time trends. The interaction between being female and of college-entering-age after the year of the policy change is positive and significant. This effect remains stable when year of birth fixed effects are introduced. Then, in columns 3 and 4, differential post trends are introduced, revealing that the effect is driven primarily by the change in slope. This is expected since, as discussed above, the exposure to treatment is in fact gradual and depends on the path taken by each individual up to the time of change. We therefore treat this as our main specification and repeat it using GLS estimation in columns 5 and 6. For this purpose our unit of observation is the group of same gender individuals born in a specific year (e.g. men born in 1968), and we collapse the data to means accordingly.

To complement these results, in Table 5, we present a regression that classifies a portion of each cohort as treated based on which percentage of individuals would have not yet entered college, according to data from the 1995 Census on college entry ages. These percentages are allowed to be different for men and women, which accounts for the fact that men on average enter college later. These results again show a significant impact of being in the “treated” cohorts.

We also perform placebo test using high school completion, which should not be affected by the policy change, since we do not expect teen-aged girls to make this decision based on their fertility prospects. Figure A1 demonstrates that, as expected, there is no impact on this outcome either if we are timing the break at the same cohort as for college graduation (1971—in case there was a shock to this cohort’s educational outcomes for non-IVF reasons) or if we use those who would be entering their junior year (and thus still able to change their high school completion decision) in 1994 (1978 cohort—in case there was a shock to all educational outcomes in 1994 not driven by the IVF policy change).

Graduate Education We next examine whether more women completed graduate education following the policy change. For this outcome measure, we again use the median age of students entering that educational level to guide us, therefore treating the 1966 cohort as the first treated year.[34] The raw data is shown in Figure 6, showing again a clear increase in women’s completion relative to men starting at the cohorts who have not completed their educational decisions before they learn of expanded access to IVF. While women’s educational outcomes remain on a moderately increasing relatively stable trend in the pre-period, for men we find inexplicable low rates of graduate education for the 1954 to 1958 cohorts.[35] Nevertheless, women demonstrate an upward shift at the 1966 cohort, while men seem to follow approximately the same pre-trend (if we disregard the

[34]The median age for second degree applicants in Israel is 28.2 for women and 29.7 for men.

[35]We can speculate that these cohorts may have been entering military service during the 1973 Yom Kippur war and the period of hostility that followed, which may have impacted their long-term educational attainment, but have not found any literature indicating a reason for this decline.

Table 4: College Graduation Rates

	Dependent variable: College Education					
	DiD with GSTT		Slope-Change DiD		GLS Slope-Change DiD	
	(1)	(2)	(3)	(4)	(5)	(6)
fem × post	0.025	0.025	0.009	0.009	0.020	0.013
	(0.011)**	(0.006)***	(0.009)	(0.008)	(0.012)*	(0.009)
	[0.016]	[0.016]	[0.015]	[0.015]		
fem × post × time			0.007	0.007	0.006	0.007
			(0.002)***	(0.002)***	(0.003)**	(0.002)***
			[0.004]*	[0.004]*		
fem × time	0.003	0.003	0.002	0.002	0.002	0.002
	(0.001)***	(0.000)***	(0.001)***	(0.000)***	(0.001)**	(0.000)***
	[0.001]***	[0.001]***	[0.001]**	[0.001]**		
post × time			-0.009	-0.009	-0.001	0.002
			(0.001)***	(0.001)***	(0.003)	(0.001)*
			[0.003]***	[0.003]**		
post	-0.015		0.007		0.021	
	(0.009)		(0.006)		(0.011)*	
	[0.012]		[0.010]			
female	0.059	0.059	0.055	0.055	0.051	0.053
	(0.006)***	(0.004)***	(0.005)***	(0.004)***	(0.009)***	(0.005)***
	[0.025]**	[0.025]**	[0.026]**	[0.026]**		
time	0.005	0.004	0.006	0.006	0.004	0.004
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.001)***	(0.000)***
	[0.001]***	[0.001]***	[0.001]***	[0.001]***		
Constant	0.505	0.477	0.512	0.512	0.326	0.329
	(0.007)***	(0.008)***	(0.005)***	(0.005)***	(0.009)***	(0.003)***
	[0.026]***	[0.031]***	[0.026]***	[0.026]***		
FEs		YES		YES		YES
Observations	173790	173790	173790	173790	54	54
R-Squared	0.108	0.109	0.109	0.109		

Notes: Columns 1–4: Ordinary least-squares difference-in-differences regression using micro data, including controls for religiosity and parents' origin. Robust standard errors clustered at the gender × year level in parentheses; robust standard errors clustered at the gender × geography level in square brackets. Columns 5–6: Generalized least squares regression with data collapsed to the gender-year of birth level. Robust standard errors that allow for cross-sectional correlation and for serial correlation with panel specific correlation parameter, in parentheses. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

*** p<0.01, ** p<0.05, * p<0.1

Table 5: College Education by Percent of Cohort Treated

	Dependent variable: College Education			
	DiD with GSTT		GLS DiD	
	(1)	(2)	(3)	(4)
fem × % treated	0.028	0.040	0.026	0.044
	(0.017)*	(0.010)***	(0.024)	(0.017)***
	[0.027]	[0.028]		
fem × time	0.003	0.002	0.003	0.002
	(0.001)***	(0.000)***	(0.001)***	(0.001)**
	[0.001]**	[0.001]*		
% cohort treated	-0.026	0.014	0.056	0.005
	(0.013)*	(0.023)	(0.019)***	(0.034)
	[0.020]	[0.035]		
female	0.057	0.052	0.065	0.050
	(0.010)***	(0.006)***	(0.016)***	(0.012)***
	[0.025]**	[0.024]**		
time	0.005	0.004	0.002	0.004
	(0.001)***	(0.001)***	(0.001)**	(0.001)***
	[0.001]***	[0.001]**		
Constant	0.513	0.465	0.300	0.339
	(0.009)***	(0.020)***	(0.013)***	(0.027)***
	[0.026]***	[0.038]***		
FEs		YES		YES
Observations	173790	173790	54	54
R-Squared	0.108	0.109		

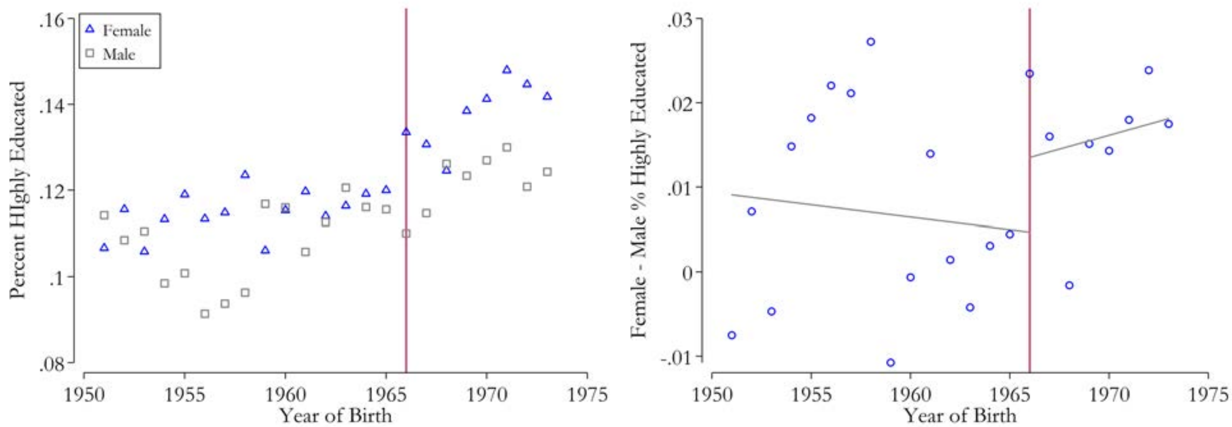
Notes: Independent variable is the percent of the cohort—defined separately for males and females—that has yet to enter college in 1994, based on data from the 1995 census on the age distribution of college freshmen. Columns 1–2: Ordinary least-squares difference-in-differences regression using micro data, including controls for religiosity and parents' origin. Robust standard errors clustered at the gender × year level in parentheses; robust standard errors clustered at the gender × geography level in square brackets. Columns 3–4: Generalized least squares regression with data collapsed to the gender-year of birth level. Robust standard errors that allow for correlation within and across time periods in parentheses. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

*** p<0.01, ** p<0.05, * p<0.1

aforementioned irregularity).

Figure 6: Difference in Percentage of Highly Educated Female and Male

(a) Separate (b) Difference



Notes: Figure (a) shows average graduate school attainment for women and men by birth cohort. Figure (b) presents the difference in graduate school attainment between women and men, as well as fitted lines for the pre (1951-1965 birth cohorts) and post (1971-1973 birth cohorts) periods. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

This is confirmed by the findings presented in Table 6. Columns 1–2 show estimates for the simple D-i-D specification with gender-specific time trends. Women in the “treated” cohorts are significantly more likely to complete graduate degrees than before. When allowing for a discontinuous slope change, the main effect is relatively stable and the slope change, while positive, is not significant. This can be due to the fact that we use relatively few post period cohorts to avoid censoring, in addition to the unstable trend exhibited for men in the pre period (as discussed above) and to graduate education generally being noisier than college. Nevertheless, the positive effect remains stable when year-of-birth fixed effects are included, and the GLS specification confirms the magnitude of the results, although it is only marginally significant in column 5, and not significant in column 6.

We also look at whether rates of women gaining graduate education *conditional* on obtaining college education have increased, in table 7. In doing this, we seek to understand whether graduate education has increased as a natural consequence of the increase in college education, or whether there has been an increase in graduate education over and above the mechanical impact of increasing the pool of college graduates. The estimated change ranges between 5 and 6 percentage points, when conditioning on college completion, which amounts to approximately 13% increase relative to the baseline level (at the year prior to the change). These results are substantially larger and more significant for all specifications compared to the unconditional graduate education estimates. Moreover, the magnitude of the effect is larger than the effect we see on college graduation, both in

Table 6: Rates of Graduate Education

	Dependent variable: Graduate Education					
	DiD with GSTT		Slope-Change DiD		GLS Slope-Change DiD	
	(1)	(2)	(3)	(4)	(5)	(6)
fem × post	0.018	0.018	0.016	0.016	0.016	0.014
	(0.007)**	(0.005)***	(0.006)**	(0.005)***	(0.009)*	(0.009)
	[0.009]**	[0.009]**	[0.008]*	[0.008]*		
fem × post × time			0.001	0.001	0.001	0.001
			(0.001)	(0.001)	(0.002)	(0.002)
			[0.002]	[0.002]		
fem × time	-0.000	-0.000	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
	[0.001]	[0.001]	[0.001]	[0.001]		
post × time			-0.003	-0.003	0.001	0.001
			(0.001)***	(0.001)***	(0.001)	(0.001)
			[0.002]**	[0.002]		
post	-0.014		-0.008		-0.003	
	(0.005)***		(0.005)*		(0.007)	
	[0.006]**		[0.006]			
female	0.000	0.000	-0.001	-0.001	-0.001	-0.000
	(0.005)	(0.003)	(0.004)	(0.003)	(0.007)	(0.006)
	[0.012]	[0.012]	[0.011]	[0.011]		
time	0.002	0.001	0.003	0.002	0.001	0.001
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.001)*	(0.000)
	[0.001]***	[0.000]**	[0.001]***	[0.001]***		
Constant	0.218	0.199	0.223	0.212	0.117	0.112
	(0.005)***	(0.004)***	(0.004)***	(0.006)***	(0.005)***	(0.004)***
	[0.015]***	[0.016]***	[0.014]***	[0.013]***		
FEs		YES		YES		YES
Observations	138953	138953	138953	138953	46	46
R-Squared	0.0463	0.0465	0.0464	0.0465		

Notes: Columns 1–4: Ordinary least-squares difference-in-differences regression using micro data, including controls for religiosity and parents' origin. Robust standard errors clustered at the gender × year level in parentheses; robust standard errors clustered at the gender × geography level in square brackets. Columns 5–6: Generalized least squares regression with data collapsed to the gender-year of birth level. Robust standard errors that allow for cross-sectional correlation and for serial correlation with panel specific correlation parameter, in parentheses. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

*** p<0.01, ** p<0.05, * p<0.1

percentage points and in percentage relative to baseline level. This supports our main hypothesis that extended later-life fertility for women drives the observed shifts, since decisions on graduate education are made at an older age when expected fertility plays a much more important role.

Table 7: Conditional Rates of Graduate Education

	Dependent variable: Graduate Education | College					
	DiD with GSTT		Slope-Change DiD		GLS Slope-Change DiD	
	(1)	(2)	(3)	(4)	(5)	(6)
fem × post	0.062	0.062	0.056	0.056	0.050	0.049
	(0.021)***	(0.014)***	(0.018)***	(0.015)***	(0.026)*	(0.027)*
	[0.022]***	[0.021]***	[0.022]**	[0.021]***		
fem × post × time			0.003	0.003	0.003	0.003
			(0.003)	(0.002)	(0.005)	(0.005)
			[0.004]	[0.004]		
fem × time	-0.005	-0.004	-0.005	-0.005	-0.005	-0.005
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.002)**	(0.002)**
	[0.002]***	[0.002]***	[0.002]***	[0.002]***		
post × time			-0.009	-0.005	-0.006	0.003
			(0.002)***	(0.002)*	(0.004)	(0.003)
			[0.004]**	[0.005]		
post	-0.054		-0.039		-0.031	
	(0.017)***		(0.015)**		(0.022)	
	[0.017]***		[0.016]**			
female	-0.055	-0.055	-0.059	-0.060	-0.060	-0.059
	(0.015)***	(0.010)***	(0.012)***	(0.010)***	(0.018)***	(0.018)***
	[0.014]***	[0.014]***	[0.014]***	[0.014]***		
time	0.001	-0.002	0.003	-0.001	0.002	-0.004
	(0.001)	(0.001)***	(0.001)**	(0.001)	(0.002)	(0.001)***
	[0.001]	[0.001]*	[0.002]**	[0.002]		
Constant	0.463	0.413	0.476	0.431	0.417	0.352
	(0.014)***	(0.007)***	(0.012)***	(0.012)***	(0.015)***	(0.010)***
	[0.012]***	[0.015]***	[0.012]***	[0.015]***		
FEs		YES		YES		YES
Observations	45609	45609	45609	45609	46	46
R-Squared	0.0128	0.0136	0.0132	0.0136		

Notes: Columns 1–4: Ordinary least-squares difference-in-differences regression using micro data, including controls for religiosity and parents' origin. Robust standard errors clustered at the gender × year level in parentheses; robust standard errors clustered at the gender × geography level in square brackets. Columns 5–6: Generalized least squares regression with data collapsed to the gender-year of birth level. Robust standard errors that allow for cross-sectional correlation and for serial correlation with panel specific correlation parameter, in parentheses. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

*** p<0.01, ** p<0.05, * p<0.1

Reevaluating education outcomes using repeated-cross-section data With all educational outcomes, one may be concerned about data censoring, since we use data collected in a single year and therefore compare individuals of different ages over time. To minimize this problem, the youngest cohort we use in our estimation is 31 at the census year, as discussed in section 3.2 above. In this section, we use a different data set comprised of annually repeated cross-sections, the Israeli Labor Force Survey (LFS), to verify that censoring is not what drives the result. This data allows us to compare individuals of the same age and old enough to complete their higher education of all levels, before and after the policy change. Because of the limited years of availability, and small sample size, we do not use this data as our main source.[36]

We use two-year age groups to increase the number of observations per year and decrease variation (although the sample remains quite small). We choose ages to be high enough so we can be confident that there is minimal censoring due to ongoing education, but also not "too old" so we can follow what happens to this age group for several years after 1994.[37]

Figure 7 presents the percentage of college graduates in each cohort, separately for men and for women on the left hand side and differences on the right. We clearly see that while men stay on the same moderately increasing time trend, women's rate of college completion sharply increases starting with the 1971 cohort, similarly to what we see in the Census data. This result refutes the possibility that the results we presented above are the result of data censoring which is more severe for men. The same analysis is presented in Figure 8 for graduate level education. Interestingly, there are two 'jumps' for women, the first for the 1966 cohort and the second for the 1971 cohort. It seems reasonable that the first increase is driven by women who had already completed college when the policy was introduced and due to the policy faced a decreased cost of attending graduate education. The second increase correlates with the increase in college attainment and is at least partially driven by the higher rates of women who are college graduates and can actually consider post college education.

This data set is sampled and assembled completely differently from our principal data, and yet shows remarkably similar results, thus providing additional evidence that there was a differential increase in women's investment in higher education, starting with the cohorts who had the opportunity to enter either college or graduate school in 1994.

4.2 Marriage Market Equilibrium

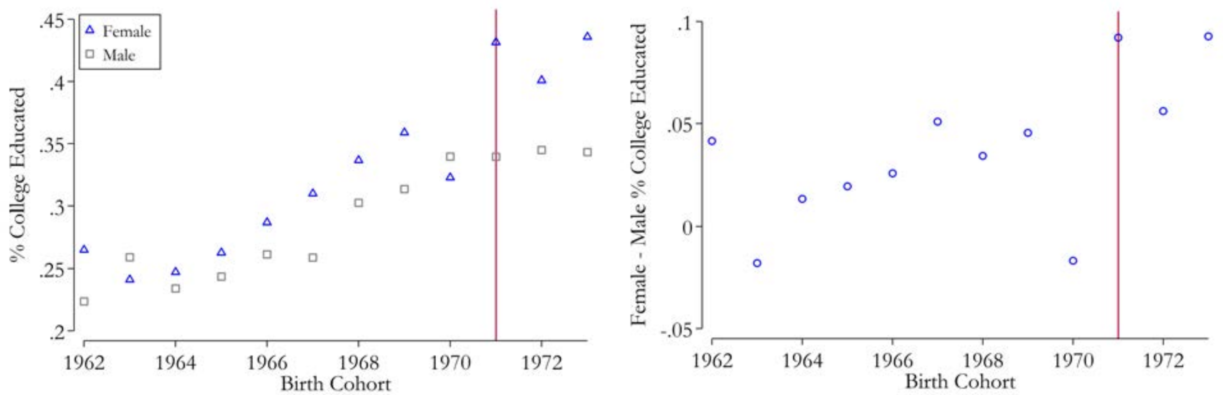
The additional reproductive years afforded by access to assisted reproduction technologies may have impacted not only women's decisions, but also men's marriage choices. Low (2016) shows that men respond to prospective mates' expected fertility when choosing a partner, trading off between so-

[36]We utilize The Israeli Annual Labor Force Survey for the years 2001 to 2011. This sample is representative of the population, but much smaller than the Census sample. The LFS is a longitudinal survey following the size and evolution of the labor force in Israel at the household level. Due to the nature of this data, households may have been surveyed multiple times within each survey year, and so we restrict our data to focus on unique observations (eliminating ambiguous cases that have missing IDs and cannot be verified as unique). We identify a total of 976,322 unique observations from 1970 to 2011, approximately 23,246 observations per year.

[37]It should be noted that our results are not sensitive to this choice and the same pattern appears for a variety of age ranges.

Figure 7: LFS Percentage of College Educated Female and Male, For 38-39 Year-Old Cohort

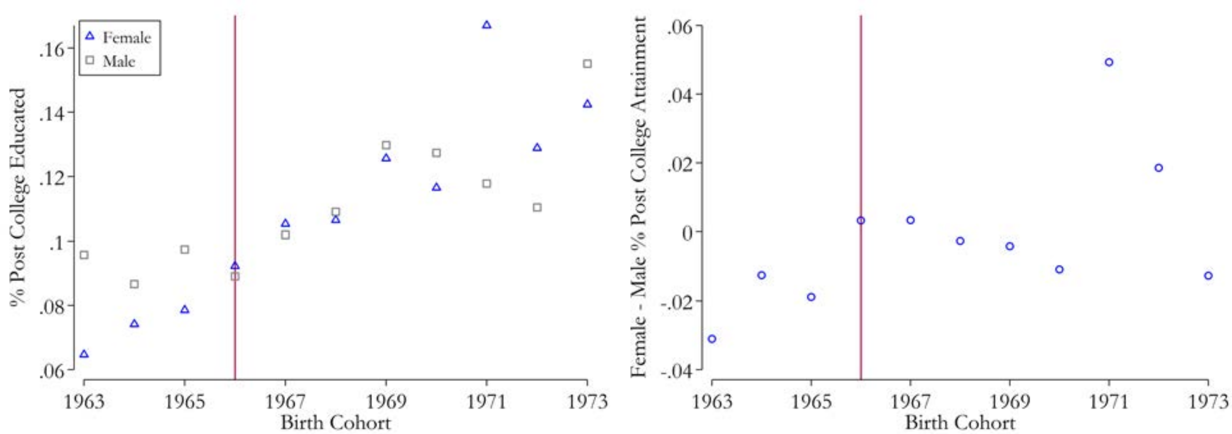
(a) Separate　　　　(b) Difference



Notes: Figure (a) presents the fraction of men and women with college education by birth cohort. Figure (b) presents the difference between women and men. Data from the Labor Force Survey 2001-2011, restricted to Israeli-born Jews.

Figure 8: LFS Percentage of Graduate Educated Female and Male, For 38-39 Year-Old Cohort

(a) Separate　　　　(b) Difference



Notes: Figure (a) presents the fraction of men and women with graduate education by birth cohort. Figure (b) presents the difference between women and men. Data from the Labor Force Survey 2001-2011, restricted to Israeli-born Jews..

called "reproductive capital" and more traditional human capital traits like income and education. As a result, women who are high-earning, but older, may marry poorer men than lower-earning, but younger, women. Because the increase in access to IVF technology lessens the perceived fertility cost of waiting to marry, "high-quality" men may have been more willing to marry older women following the policy change. If this is the case, we can expect equilibrium matching to adjust so that these women will match with higher quality partners.

We test this by examining the spousal quality of women who marry older versus younger before and after the policy change. If women's reproductive fitness is taken into account by men, we would expect the "spousal quality penalty" to older women to lessen once access to IVF expands.

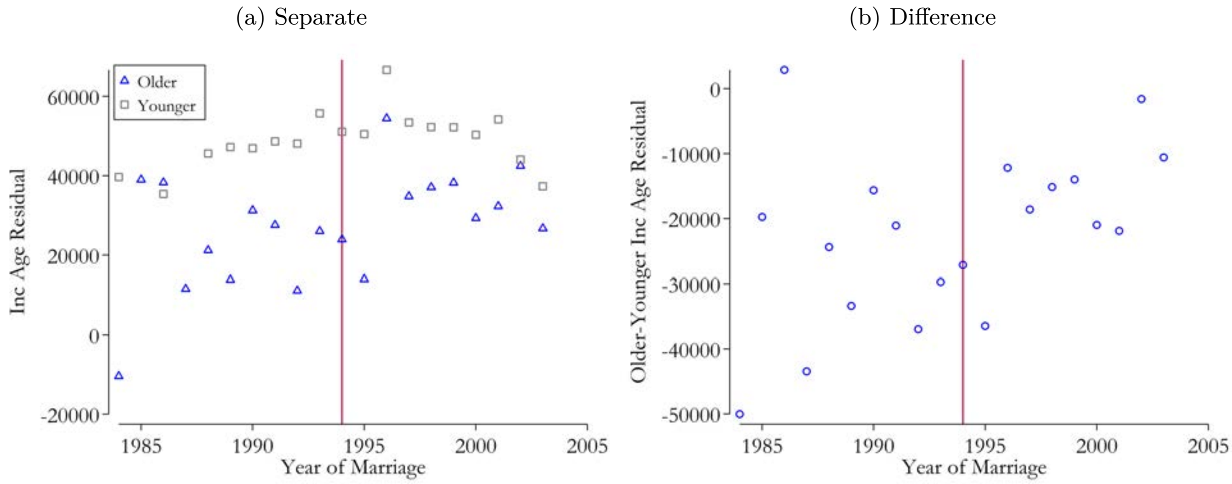
We restrict our sample to married women who were between 25 and 34 at marriage. Then, we compare spousal quality, measured in a variety of ways, for women who were between 25 and 29 at the time of marriage versus women who were between 30 and 34 at the time of marriage. The age restriction ensures that we compare relatively similar groups of women. At the same time, placing the cutoff at the age of 30, helps us identify the different response to the change in expected fecundity, which becomes significantly more relevant when a woman hits her thirties.

Our main proxy for spousal quality is husband's income, as it is well established that income is an important quality that male spouses bring to the relationship (see, for example, Fisman et al. (2006)). To control for age effects, since we use cross-sectional data and since women who marry at different ages may have spouses of different ages as well, we regress husband's income on a flexible polynomial in age, and take the residual as our outcome variable. Since we only have spousal income data for the current spouse, in practice, we cannot conduct this analysis for first marriages only. Therefore, we are missing data on women who are divorced or widowed before the census year, which may become more likely with more distant years of marriage. This data structure also means that for very young or very old spouses, income may provide a distorted measure of quality, since they may still be completing school and career training, or already retired, accordingly. To minimize these two potential problems, we restrict our analysis to the years 1984 to 2003, and verify our results hold for an even shorter period between 1988 and 1999. In addition, we use college education as an alternative measure for spousal quality, and further control for spousal age by adding spouse year-of-birth fixed effects (these results are presented in appendix table A2).

The raw data for this analysis is shown in Figure 9. Starting in 1996, two years after the policy change was announced, we observe a distinct shift upward in spousal income for women who marry over 30, relative to the same measure for women who marry younger, which remains very stable over time. This clear "break" in the series of spousal income for "older" brides, indicates that the marriage market equilibrium may have been impacted by the IVF policy change. The fact that this break appears with a lag may be attributed to men's perceptions taking longer to update or simply to the fact that reaching this new matching-equilibrium takes time.

Table 8 presents the results from a regression of income-age-residuals on marrying older, before and after the policy change. The results confirm that there is a significant "penalty" in terms of spousal income for women who choose to marry over thirty, and that this penalty significantly

Figure 9: Income-Age-Residual, 1984 - 2003



Notes: This figure compares current spousal income for women who married at age 30-34 to women who married at age 25-29. In order to control for age effects, the measure of spousal income used is the residual from a regression of income on a flexible polynomial in age. We use a narrower time range in these graphs to prevent censoring from individuals either still being students, for younger cohorts, or entering retirement, for older cohorts. Figure (a) shows average income-age-residuals separately by age group, by year of marriage. Figure (b) presents the differences in spousal income-age-residual, as well as fitted lines for the pre (1984-1993) and post (1994-2003) periods. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

decreases in the post period. Depending on the specification, we observe a reduction of 60 to 80 percent in the “older marriage penalty” for women. This change is significant with both types of clustering, and in the GLS specification. As previously mentioned, we do not include a specification that allows for a change in slope for this outcome measure, due to the narrow band of years and the lagged response. To further demonstrate the robustness of this shift, we narrow the window of years to a twelve-year period from 1988-1999. As shown in the bottom part of table 8, the estimated effect remains stable and significant for all specifications.

In the appendix, we conduct three further robustness checks for the spousal quality results, shown in Table A2. First, in columns 1–2, we repeat the income-age-residual regressions controlling for wife’s characteristics, to ensure that the increase spousal income is not a result of the improved quality of the older brides themselves (in terms of income and education). The latter is expected given our results regarding the effect of IVF funding on women’s higher education. Second, in columns 3–4, we try to control more flexibly for spousal age effects that may be time varying by using raw spousal income as our outcome variable, but adding spouse’s year of birth fixed effects. This way, the effect of older brides somehow getting a better cohort of men in later years is removed, giving a clearer measure as to whether spousal quality conditional on cohort has improved. Finally, in columns 5–6, we use spousal college education as our quality outcome variable, which should not

Table 8: Spousal Income (residual of regression of income on flexible age polynomial)

	Dependent variable: Income Age Residual, 1984-2003			
	DiD		GLS DiD	
	(1)	(2)	(3)	(4)
older × post	25159.599	25287.515	18292.693	18237.947
	(7745.650)***	(5382.890)***	(4626.508)***	(5051.319)***
	[8322.069]***	[8332.364]***		
post	7306.007		7006.300	
	(7105.614)		(6040.773)	
	[6495.848]			
married older	-30660.774	-30418.114	-30751.373	-30815.420
	(5742.011)***	(4644.425)***	(3734.606)***	(4091.786)***
	[12213.127]**	[12151.839]**		
time	-1727.504	-491.605	-691.452	-326.156
	(681.360)**	(474.453)	(505.656)	(56.732)***
	[517.630]***	[642.732]		
Constant	89697.866	85246.577	48742.545	39789.584
	(7923.620)***	(6414.770)***	(3630.797)***	(529.149)***
	[11845.781]***	[14024.894]***		
FEs		YES		YES
Observations	19458	19458	40	40
R-Squared	0.0433	0.0453		
	Dependent variable: Income Age Residual, 1988-1999			
	DiD		GLS DiD	
	(1)	(2)	(3)	(4)
older × post	21088.805	20919.673	15960.725	15771.478
	(7771.923)**	(6595.255)***	(4399.008)***	(4421.504)***
	[9805.351]**	[9812.762]**		
post	-6682.513		-3893.632	
	(5198.565)		(6087.773)	
	[9817.846]			
married older	-29153.167	-29223.528	-32294.700	-32351.989
	(5567.066)***	(5448.469)***	(3339.732)***	(3354.346)***
	[15404.833]*	[15357.801]*		
time	536.751	-243.731	1209.370	876.922
	(771.530)	(283.401)	(855.857)	(80.761)***
	[1350.522]	[1245.313]		
Constant	108863.268	101709.679	56543.669	52411.072
	(5428.510)***	(4420.773)***	(3737.098)***	(423.268)***
	[12835.829]***	[11534.700]***		
FEs		YES		YES
Observations	11048	11048	24	24
R-Squared	0.0499	0.0506		

Notes: Columns 1–2: Ordinary least-squares difference-in-differences regression using micro data, including controls for religiosity and parents' origin. Robust standard errors clustered at the age group × year level in parentheses; robust standard errors clustered at the age group × geography level in square brackets. Columns 3–4: Generalized least squares regression with data collapsed to the age group-year of marriage level. Robust standard errors that allow for cross-sectional correlation and for serial correlation with panel specific correlation parameter, in parentheses. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

be time-varying once a degree is obtained, and thus is less likely to be skewed by the retrospective analysis. All three of these alternative approaches to measuring spousal quality provide positive, significant results for the change in spousal quality for brides over thirty. In terms of magnitude, these estimates are on the high end of our main results, indicating that 80-100 percent of the "older marriage penalty" dissipated in the post period.

For this outcome measure, we can also compare the older-versus-younger results to the income gaps of men's spouses based on their age at marriage, using a triple-difference specification. In this specification, the spousal income for women who marry older versus younger, before and after the policy change are compared to the same metrics for older versus younger *men* before and after the policy change. The results, presented in Table 9, show that not only do older women's marriage outcomes (in terms of spousal quality) improve relative to younger women following the policy change, but they also improve relative to the change in older *men*'s spousal quality.

Together, these results show that women delayed marriage and made greater educational investments after the policy change. At the same time, women who delayed marriage were penalized less on the marriage market for their older age at marriage. These profound effects are consistent with women believing the policy (and the technology that they learned about through the policy change) provided some insurance against age-related infertility, altering both their decisions and the decisions of men who may have updated their beliefs regarding older women's fertility prospects. The combination of these results and effects on different outcomes strengthen our conclusion that IVF availability was responsible for each individual change.

4.3 Robustness Checks and Alternative Explanations

In this section, we perform robustness checks of our results and examine some potential alternative explanations for our findings.

Permutation approach to standard errors In micro-data difference-in-differences studies, one may be concerned that a high degree of intra-group correlation or correlation across time periods is driving the significance of the results. We have addressed this issue by providing two alternative ways to cluster the standard errors, as well as GLS estimates that are collapsed to the year-group level and allow for correlation within groups and across years. As an additional check that our estimates are large relative to the true variation in the data, we also perform two types of permutation analyses on our coefficients for the DiD specifications (since these have only a single coefficient of interest, unlike the regressions that allow for a slope change), presented in Figures A4 – A7.

For each outcome, we first perform a permutation test that respects the potential serial correlation in the data, by implementing a regression like our "column 1" specification for only ten years of data, five years pre and five years post, with 1994 (or the corresponding school entry cohort)

Table 9: Spousal Income (Residual of Income Regression on Age Polynomial), Triple Difference

	Dependent variable: Income Age Residual, Triple Difference			
	DiD		GLS DiD	
	(1)	(2)	(3)	(4)
older × post × fem	19264.231	19124.935	20944.459	19300.130
	(8249.682)**	(7710.250)**	(7250.807)***	(7510.576)**
	[9087.631]**	[9125.961]**	[7250.807]***	[7510.576]**
older × post	3176.984	3332.265	-802.185	-1064.562
	(3562.150)	(3694.837)	(3034.490)	(3078.584)
	[3032.765]	[3100.444]	[3034.490]	[3078.584]
older × fem	-27370.021	-27232.458	-29126.677	-27887.224
	(5561.731)***	(5698.597)***	(5901.454)***	(6131.403)***
	[13584.554]**	[13573.406]**	[5901.454]***	[6131.403]***
fem × post	-20601.859	-20427.341	-21709.236	-19997.046
	(4371.866)***	(3660.908)***	(6399.644)***	(6505.731)***
	[4000.092]***	[4001.056]***	[6399.644]***	[6505.731]***
post	11813.784		8482.344	
	(4589.010)**		(2906.153)***	
	[2934.627]***		[2906.153]***	
female	63610.949	63523.157	65698.758	64285.282
	(2527.527)***	(2107.570)***	(5277.540)***	(5384.216)***
	[8973.068]***	[8951.417]***	[5277.540]***	[5384.216]***
married older	-3487.481	-3467.365	-2148.070	-2045.975
	(2719.709)	(2027.918)*	(2399.755)	(2436.960)
	[3748.455]	[3766.264]	[2399.755]	[2436.960]
time	187.028	814.182	1062.643	1036.956
	(410.960)	(318.883)**	(194.471)***	(132.182)***
	[309.727]	[289.916]***	[194.471]***	[132.182]***
Constant	19171.227	25258.766	-7058.913	-3759.558
	(4255.123)***	(3731.286)***	(1968.865)***	(1284.212)***
	[3607.443]***	[4429.847]***	[1968.865]***	[1284.212]***
FEs		YES		YES
Observations	47779	47779	80	80
R-Squared	0.0660	0.0666		

Notes: Columns 1–2: Ordinary least-squares triple-difference regression using micro data, including controls for religiosity and parents' origin. Robust standard errors clustered at the gender × age group × year level in parentheses; robust standard errors clustered at the age gender × group × geography level in square brackets. Columns 3–4: Generalized least squares regression with data collapsed to the gender-age group-year of marriage level. Robust standard errors that allow for cross-sectional correlation and for serial correlation with panel specific correlation parameter, in parentheses. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

*** p<0.01, ** p<0.05, * p<0.1

as the treatment year. We then compare this coefficient to the coefficient obtained from taking each possible sequential interval of ten years within our study period, and each corresponding false "treatment" year in the middle of the interval. This test does not yield a normal distribution of coefficients, as there are a limited number of ten-year intervals in the study period. Our true effect is larger than the effect of any other treatment year for age at first marriage and graduate education. For college education only one of the coefficients is (slightly) larger (which amounts to approximately 6% of the values). For spousal income, we see higher estimated changes for the 3 years following our "true" treatment year, which corresponds to the "delayed response" of the marriage market equilibrium (as discussed above).

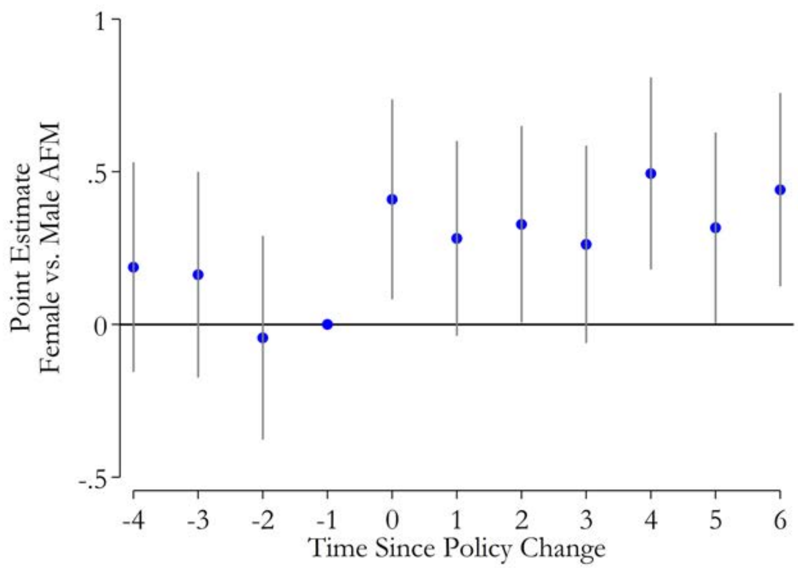
We then perform a more standard permutation test, where we randomly draw a number of years equal to our true number of treated years from the entire study period, and run a "column 1" regression with these randomly selected years as the "treatment" period (for an example of this approach, see Agarwal et al (2015)). This approach does not respect the underlying serial correlation in the data, since the years are drawn randomly, but does account for intra-group correlation or other non-standard error structures. We perform 1,000 such random draws, and compare our true treatment coefficient to the resulting normal distributions. Our true effect is outside of the curve for every outcome measure except for graduate education, for which the true effect is in the far right tail, with less than 5% of the values being above it.

Event Study analysis The next potential confounding factor we explore is that long term time trends may be responsible for the effects we see. This is already partially addressed by the inclusion of group-specific time trends in our regressions. However, to further address this possibility, we perform an event study analysis (also known as dynamic lags analysis), to pinpoint the timing of the changes we observe. We do this for our main outcome measures: college education, graduate education, age at first marriage, and spousal income for women who marry when older.

The event study graphs depicted in Figures 10 – 13 are created by regressing our key outcome variable on a series of dummies for each year, interacted with gender or age respectively (gender for educational outcomes, age at marriage for spousal income). The coefficients graphed represent the effect of each time period, controlling for all other time periods. The coefficient on the lag just before the policy change is normalized to zero, so that subsequent effects show the relative difference in the affected group's outcomes compared to the period just before the policy change. Note that the time periods we are looking at are quite small relative to our other graphs, as we are zeroing in on 4 years before and 6 years after the policy change only. All event studies show that there are no significant pre trends driving our results, and that outcomes were relatively flat in the years immediately preceding the policy change. The event studies for age at first marriage, college education, and graduate education show that results become significant after the policy change, either immediately or over time. Appendix Figures A8 – A11 show an alternate format of event study, a “distributed lag” analysis, where coefficients for being born at or above a certain year (rather than *in* a certain year) are graphed, controlling for all other years. This, essentially, measures the permanent shock that occurs in each year to the outcome variables.

Figure 10 shows that prior to the policy change the difference between men and women’s age at first marriage was relatively constant, showing no apparent pre-trend. Then, in the year of the policy change, there is a large and permanent change to subsequent outcomes. Figure A8 confirms this result by showing a significant and positive coefficient at time zero (which refers to 1994 or onwards).

Figure 10: Event Study: Age of First Marriage

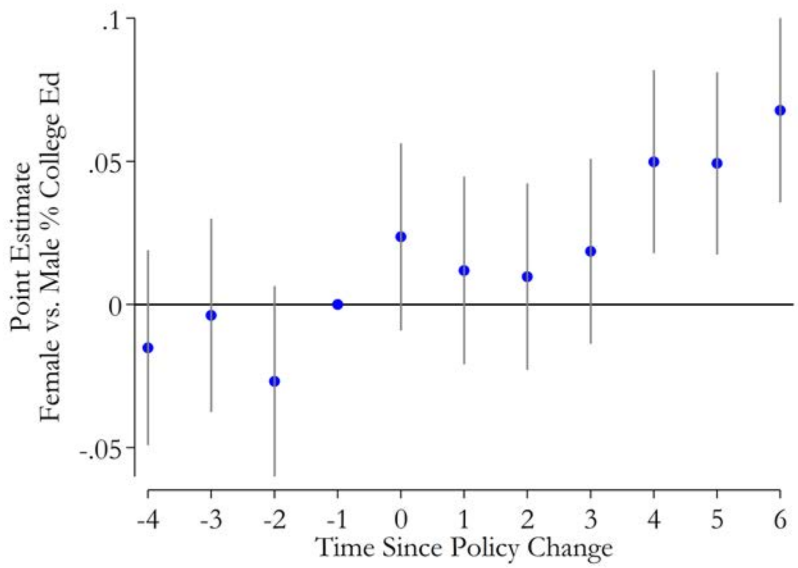


Notes: The figure presents point estimates and confidence intervals for coefficients on a series of dummy variables for being married in the specified year interacted with a dummy variable for female. The outcome variable is age at first marriage. 1994 is the time of the policy change, noted as time 0 on the x-axis.

For the education event studies, we cannot expect that the impact will necessarily be isolated to the “first treated” cohort, since preceding cohorts may have been exposed (at least partially). Indeed, figures 11 shows that the change appears to happen more gradually, only becoming significant at the end of the period. However, part of the reason for this less significant outcome is that the 1970 cohort, immediately preceding the first “treated” cohort of 1971, was also partially treated, as some of these individuals were still young enough to alter their college-going plans. This is confirmed by the distributed lags analysis in Figure A9, where *both* the 1970 and the 1971 cohorts (years -1 and 0) appear to experience a permanent positive shock of approximately equal magnitude.

Figure 12 shows that, in the case of graduate education, the effect is more clearly isolated in the “first treated” cohort, 1966 birth year, remaining permanently high, with the exception of one coefficient, after that point. Figure A10 repeats this finding.

Figure 11: Event Study: College Education

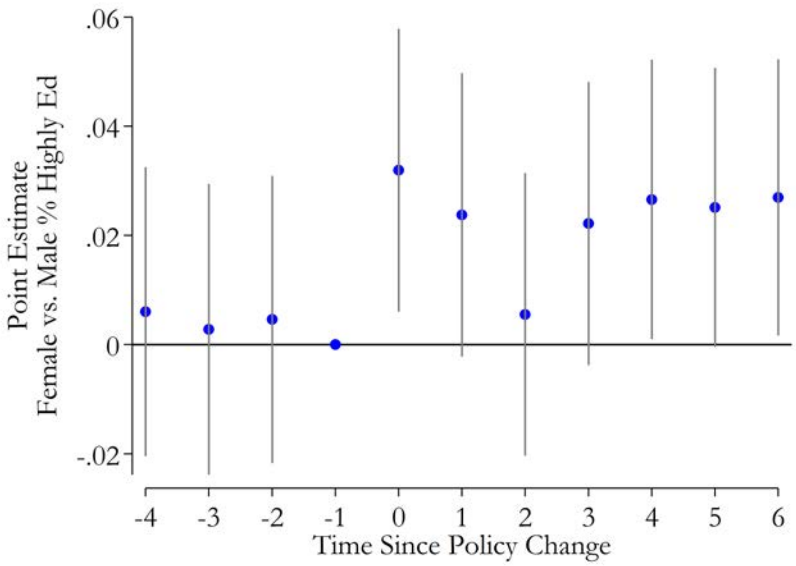


Notes: The figure presents point estimates and confidence intervals for coefficients on a series of dummy variables for having the specified birth year interacted with a dummy variable for female. The outcome variable is an indicator for college graduation. The 1971 cohort represents the first affected cohort and hence is the “time” of the policy change, noted as time 0 on the x-axis.

Figure 13 shows an event study analysis of the change in spousal income for women who marry older versus younger. This graph shows no significant effect in the period following the policy change, although it does also affirm there do not appear to be significant pre-trends prior to 1994. The same pattern is showing in Figure A11. The weaker results in this event study may be due to men’s beliefs about the fertility of older partners taking longer to update, or since shifting the marriage market to a new equilibrium takes time.

Censoring or other data issues A different possibility is that the effect we observe is an artifact created by looking at outcomes retrospectively in cross-sectional data. We already minimize this concern as we carefully choose the years and cohorts which constitute the sample for each outcome. However, to further verify that the retrospective nature of the analysis could not create similar breaks in the data *without* a real policy effect, we use the 1995 Israeli Population Census to conduct a placebo test. We replicate our analysis for a fake “policy change” in 1981 (14 years prior to the Census year, as the real 1994 policy change is 14 years before the 2008 Census) and find no evidence of a break in age at first marriage, college education, graduate education, or spousal income (age at first marriage, if anything, shifts in the opposite direction).

Figure 12: Event Study: Graduate Education

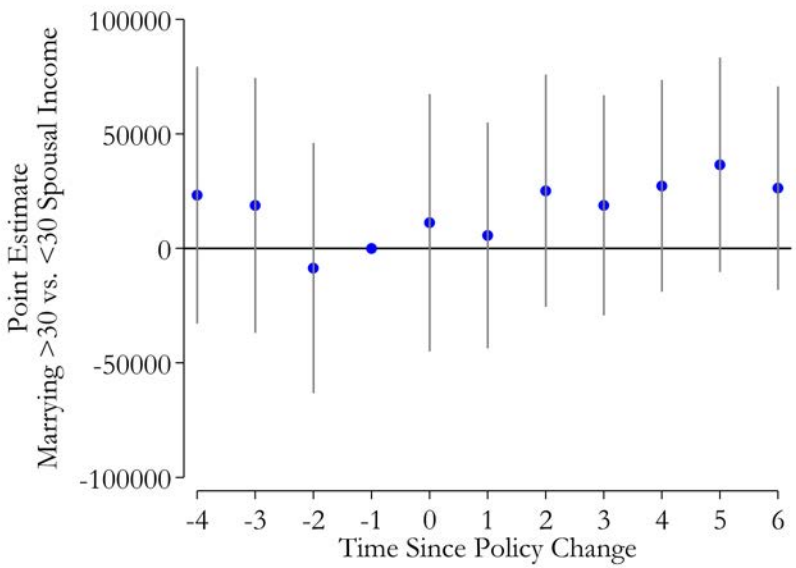


Notes: The figure presents point estimates and confidence intervals for coefficients on a series of dummy variables for having the specified birth year interacted with a dummy variable for female. The outcome variable is an indicator for post-college graduation. The 1966 cohort represents the first affected cohort and hence is the "time" of the policy change, noted as time 0 on the x-axis.

A more specific concern in this context is the possibility that men's educational outcomes were more censored than women's outcomes as we look at years closer to the present day, since men are entering and completing college education later than women. It should be noted that we already presented one method which satisfyingly deals with this obstacle; in table 5, we show that our results for college education hold when we define gradual exposure to treatment and allow men and women to differ in their level of exposure, according to the actual distribution of each group's age at college entry. Next, we further explore the robustness of our findings to this difference, by shifting data for men one year, to account for the average one year lag in male applicants' age compared to female applicants (this lag likely results from the extra year of mandated military service for men). As a result of this shift, men and women are aligned by college entry cohort, rather than by birth year. Table A1 shows these results, for both college and graduate education. The effects are slightly smaller in magnitude when accounting for this lag, yet there is still a significant increase in women's education after the policy change for all specifications.

Global shock After establishing that the break we observe is genuine, significant and timed at the year of the policy change, we turn to explore the possibility that we have misattributed the source of this dramatic change. To verify that broader international trends during the nineties are not responsible for our effects, we conduct placebo tests in the United States as well as four other

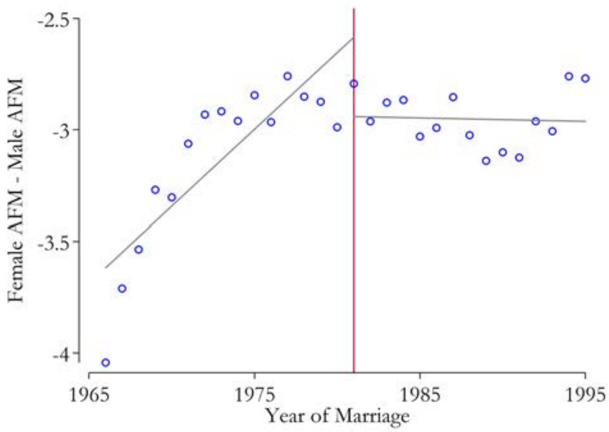
Figure 13: Event Study: Spousal Income, 1984-2003



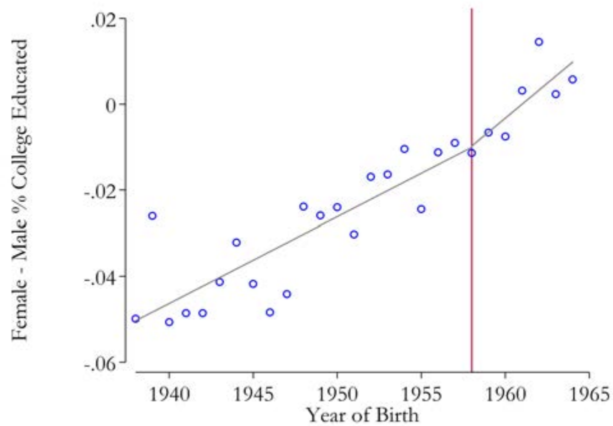
Notes: The figure presents point estimates and confidence intervals for coefficients on a series of dummy variables for being married in the specified year interacted with a dummy variable for being 30 or older at the year of marriage. The outcome variable is spousal income. 1994 is the time of the policy change, noted as time 0 on the x-axis. The sample is restricted to women with spousal matches that got married when they were older than 24 but younger than 35.

Figure 14: Placebo Test using 1995 Israeli Census

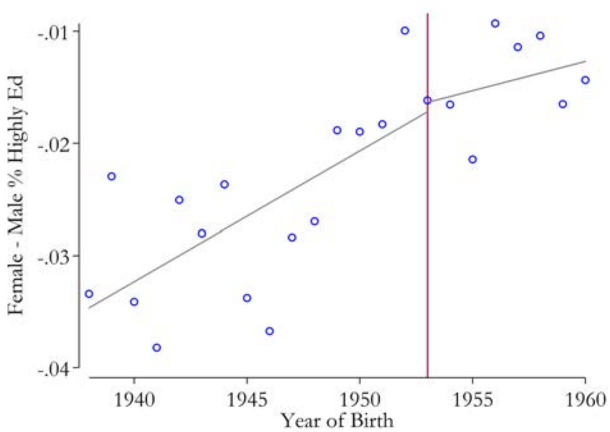
(a) Age First Marriage



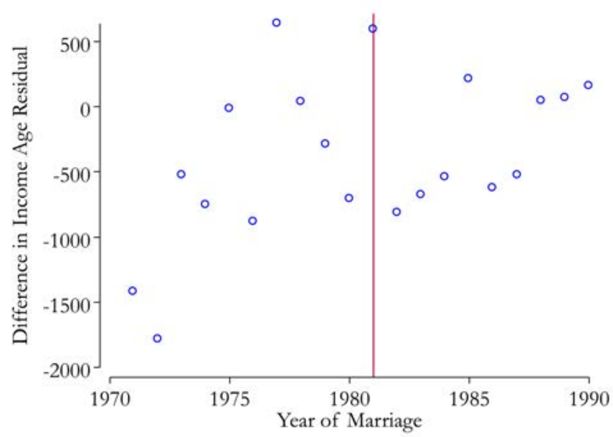
(b) College Education



(c) Graduate Education

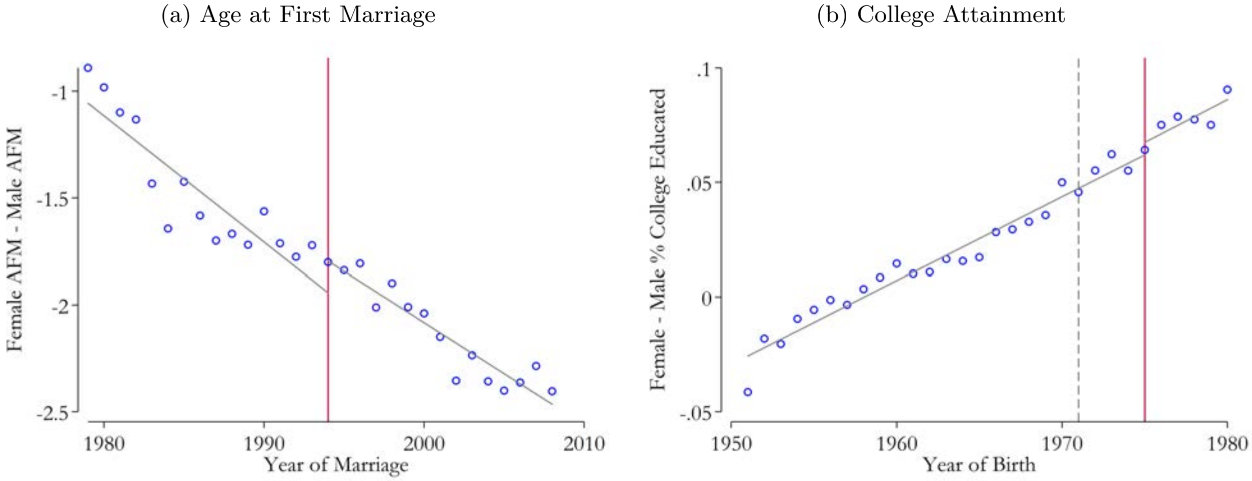


(d) Income Age Residual



Notes: Placebo test for spurious results due to retrospective analysis, using 1995 census and fictitious 1981 "policy change". Figure (a) presents the difference in men and women's age at first marriage by marriage year, figure (b) presents the difference in men and women's college attainment and figure (c) graduate attainment by birth year, and figure (d) presents the difference in spousal income-age-residual for women who marry at 30 or above versus those who marry younger. Data from 1995 Israeli population census, restricted to Israeli-born Jews.

Figure 15: United States College Attainment and Age at Marriage



Notes: Placebo test for global 1994 shift. Figure (a) shows the difference between age at first marriage for women and men, with fitted lines for the "pre" and "post" periods. Figure (b) shows the difference in college attainment between women and men, with the "treatment" year adjusted to reflect the cohort entering college in 1994 in the US. Data from the 2010 American Community Survey.

countries with similar GDP per capita to Israel, and Census data availability. Results for college education in four "comparable" countries are shown in Figure A2. The United States American Community Survey also contains information on marriage age, which allows us to look at both educational and marriage outcomes, shown in Figure 15. None of these placebo tests produce positive, significant results.

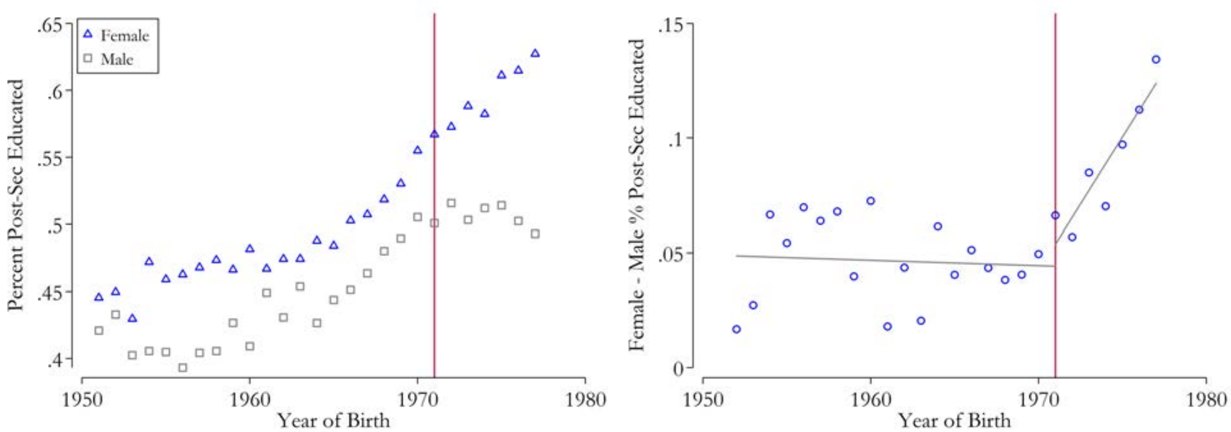
Other health expansions Another explanation of the improvement in marriage outcomes for older women in Israel might simply be the entire health services reform that the NHI law provided. Better health services can make age less important, if we believe that in the marriage market age is a proxy for health in general, rather than just fertility. However, decisions on education and marriage age should not be affected by the expansion of health services, especially since those decisions are made by young people who value those services less than older people. As far as the general insurance that better health insurance provides, there is no reason to expect that a health reform that provided the same benefits for all would have a gender divergent effect. Moreover, if anything, better public provision of health services could discourage educational and career investments, since health benefits will be provided regardless of future earnings.

Higher-education reform (and other changes affecting all women) Finally, we consider the higher education reform in Israel as an alternative explanation. This reform that was rolled out throughout the eighties and nineties, overlapping with our years of interest. Prior to the reform only universities could grant Israeli academic degrees. Starting in the seventies, colleges gradually began to receive permission to grant academic degrees equivalent to the ones given by universities. This process accelerated during the eighties and early nineties, culminating in an official and comprehensive plan for the development of academic colleges. In the decade between 1992 and 2002 the number of students in academic programs approximately doubled (the effect of the reform was already apparent in the early nineties, but really started to build up in 1997-1998 (Volanski, 2005; Bernstein, 2002)).

To verify that our results are not due to an "in name only" change in the degree individuals received, we graph the percentage of any post secondary education graduates, which will include those whose degree status would have been switched into the academic "college" category after the reform. Figure 16 shows that even if we add non-academic degrees to our analysis, we get the same trends and the same change in trend only for women. This, together with our strong results for graduate education, eases our concerns for higher education reform driving the results by re-labeling once non-academic degrees.

Figure 16: Difference in Percentage of Post-Secondary Education Attainment for Female and Male

(a) Separate (b) Difference



Notes: Figure (a) shows average post-secondary educational attainment for women and men by birth cohort. Figure (b) presents the difference in post-secondary attainment between women and men, as well as fitted lines for the pre (1951-1970 birth cohorts) and post (1971-1977 birth cohorts) periods. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

In addition, we find this alternative explanation to be unlikely, due to the different socio-economic classes targeted by the two reforms. At the time of the education reform, women already constituted more than 50% of undergraduate students. The main purpose of the reform was to make higher education institutions more accessible to a lower socioeconomic status population, mostly concentrated in peripheral regions (Volanski, 2005; Shavit et al., 2007), and increase higher education supply to match the rapidly increasing demand.[38] In addition, numerous studies were conducted to document the reform's consequences, none of which report a distinctive effect on women's participation in higher education (see for example Volanski (2005)). In fact, over the years that followed, the percentage of female students in colleges was actually lower than in universities.[39]

Moreover, similar reforms in other countries were not found to affect women differently than men. One example is the higher education reform in Spain, which was enacted at approximately the same years as in Israel, and did not change the trend of women's education or of women's marriage decisions (Mora, 1996).

It should also be noted that there is no reason to expect the reform to affect the way women's marriage outcomes depend on their age. We find that the previously existing penalty for older marriage practically disappears, even if we control for women's level of education. Combined together, it is hard to imagine that the driving force for all of the above described impacts is the increased supply of higher education rather than the increased availability of IVF technology, as we suggest.

Nevertheless, we further explore the possibility that the higher education reform differentially affected women, by testing our outcomes on the Arab population of Israel, which was excluded from our analysis so far. In Figure A3 we show the Arab-women versus Arab-men differences over time for our main outcomes and do not find the same effects at 1994 as for their Jewish counterparts. The tests show either no effect or, in the case of age at first marriage, the opposite effect. Note that the affected cohort for college and graduate education is adjusted to reflect the timing of Arab-Israelis entering college, which is younger due to no military service requirement. The dotted line reflects the cohort used in the main analysis, and there is no break apparent at that point, either.

As we mention in section 3.2, Arab-Israelis were less likely (if at all) to respond to the change in IVF funding for three main reasons. First, most Arab-Israelis are Muslim and Islam places more stringent restrictions on the use of *in vitro* fertilization than does Judaism.[40] In addition, the Arab population has a much lower average age at first marriage for women, a higher average birth rate and a much lower labor participation rate for women.[41]

[38] The demand increase stems from the growing rate of high-school graduates that received certificates in matriculation exams (which are needed when applying for college) (Shavit et al., 2007).

[39] The only exception is teacher's training colleges, where there is a vast majority of female students, however the academization process for those colleges took place in the early eighties. In addition, the students in these institutions constitute only a small share of the number of college students overall.

[40] For example, Islam prohibits the use of donor eggs or sperm, the former being extremely important and even crucial for women in their forties. In addition, the Israeli Jewish religious leadership very quickly addressed the innovative IVF technology and approved usage with practically no limitations, whereas other religions took longer to respond.

[41] In our baseline year 1993, for example, Arab women appear to marry 2.5 years earlier than Jewish women and the pre-trend for Arabs is positive but much more moderate (Also see Macro data on marriage age and TFR reported in footnote 27 above). Labor force participation rate in the early-mid 90s was approximately 13% for Arab-Israeli Muslim women compared to around 55% for Jewish Israeli women (based on data from the Labor Force Surveys and

Second, Arab-Israeli women were much less likely to be on the margin of large career investments in the 1990s, as average educational levels were substantially lower than in the Jewish population. At our baseline year 1993, there is a 25 percentage point difference in the rate of women's college attainment between Jews and Arabs, and a 10 percentage point difference in the same figures for graduate education (Arab women's attainment in graduate education was actually very close to zero at that time). Third, since the Arab population is not subject to obligatory military service, they tend to make decisions about career investment when they are 2-3 years younger, making fertility considerations less relevant for educational decisions.

However, Arab-Israelis were much more likely than Jewish-Israelis to be affected by the higher education reform due to lower high-school achievements (on average) and higher concentration in peripheral areas.[42] For the same reasons, Arab-Israeli women would have been targeted by, and affected by, any other government programs designed to increase women's education or "social liberation". Therefore, the same placebo test more generally helps rule out that other policies that benefited women relative to men (or any other unspecified social trends) may have been responsible for the observed changes. This provides further evidence that the changes in Jewish women's outcomes were driven by the new IVF funding policy. In the next section we further exploit the differences between the expected response of the two population groups to the policy by using Arab women as an alternative control group.

4.4 Using Arab population as control group

Our identification strategy relies on the post-1994 time-path of men's outcomes being similar to women's if it were not for the occurrence of the IVF policy change (once pre-trends and level effects have been controlled for). A threat to this identification would be a policy, or any other exogenous shock, that affected Israeli women, but not men, commencing at or around the time of the 1994 IVF policy change. The event study and QLR analyses show that such a change would need to be very precisely timed to coincide with the IVF policy change in order to produce similar results. We have already shown via a placebo test with Arab-Israelis that gender-divergent trends are unlikely to account for our result, since Arab-Israeli men and women would be likely to experience similar impacts from other policies, but are unlikely to respond to IVF availability. We now take this one step further, and replace men as our control group with Arab women, which will enable us to difference out the impact of being female following the policy change. The validity of this strategy relies on Arab-Israeli women being much less likely to use IVF due to religious restrictions, as well as less likely to be on the margin of a large career investment, but being similarly affected by other policy changes, such as general expansion of female access to education. Since the Arab

macro data reported by the Bank of Israel.

[42]This effect is described in Volanski (2005) and also in various reports issued by the Israeli council for higher education (e.g. Higher education in Israel 2014, pp. 29-31).

population may also be affected by events that are specific to their community and irrelevant to the Jewish population in Israel, this strategy would certainly be imperfect on its own. However, any bias should be uncorrelated with the bias from using men as the principal control group, and thus, the highly consistent findings presented here provide another piece of evidence for the effect of access to IVF on Jewish women in Israel. In this section we mainly focus on educational outcomes since, at least for college, the pre-trends exhibited by Jewish men and women differ, suggesting the possibility that different forces drive the educational choices of men and women. In addition, this second analysis of the educational outcomes should help eliminate the concern that higher education reform was the driver behind the boost we observe for women's college attainment.

For these results, because we want to capture any other effects that could have possibly impacted women entering college in 1994, we need to use the birth cohort of Arab women that would be entering college at the same time as Jewish women. As previously discussed, Arab women enter college earlier, most likely because they do not have a military service requirement. Military service for Jewish women is two years long, but macro data shows a three year difference in the median age of college applicants between the two populations.[43] Thus, in our figures and the following regression analysis we align the affected cohorts for the two groups to match college entry at 1994. As a result we compare Jewish women to three-years-younger Arab women. Nevertheless, all of the results hold and are qualitatively similar when we do not adjust for this difference and conduct the analysis using year of birth.

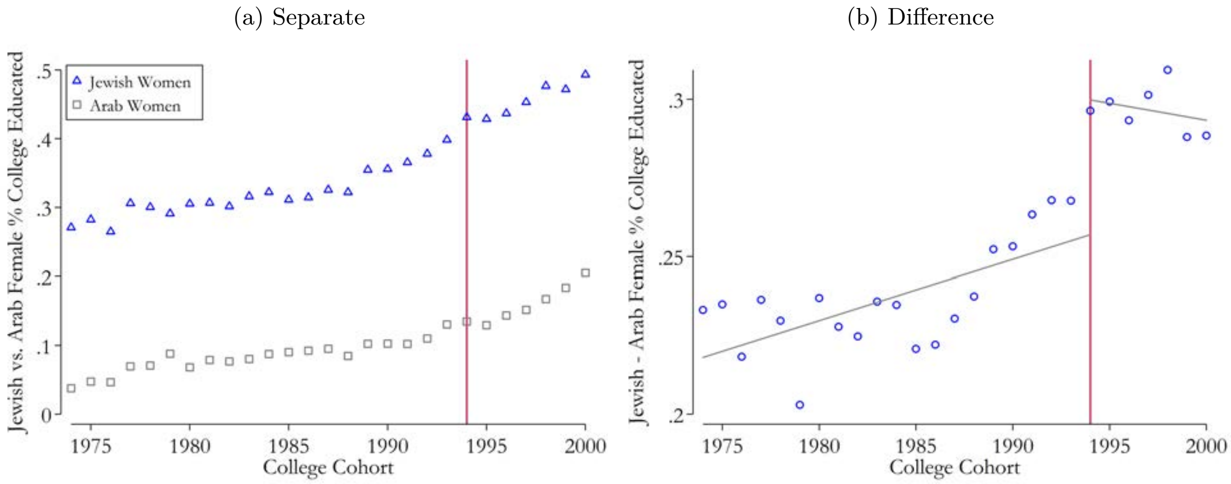
Figures 17 and 18 show the difference between Jewish and Arab women using raw data on college and graduate education completion. These figures also show these outcomes separately by population group. Although the pre-trends are not parallel, as with the male control group, a similar increase in both types of education is clearly observed at 1994. Tables 10 and 11 confirm that the results for education hold using the alternate Arab-female control group. The magnitude of the coefficients is similar to the results for the male control group. The slight negative trend seen in columns 3 and 4 is most likely because Jewish-Israeli women take longer to complete their degrees and have a higher variation in age at college entry, but could also be due to more recent increases in education among the Arab population, from policies designed to encourage education in under-served areas. However, the Jewish $\times$ post coefficient is positive and significant in all specifications.

Combining the results in this section with the ones in our main specifications establishes that the most likely cause for the observed change in Jewish women's educational choices is the increased access to IVF. Any other explanation would have to induce both a gender divergent and a religion-based divergent impact, in addition to affecting all of the outcomes we consider. Such an occurrence

[43] This is not surprising since some military occupations require a prolonged service duration and also since there is an average waiting period of 5 months between high school graduation and induction date, and then individuals may take time to pursue education at the conclusion of the assignment.

is highly unlikely.

Figure 17: Percentage of College Education by College Cohort (Arab Control Group)



Notes: The figure compares college completion rates between Jewish and Arab women over time. The cohorts are aligned based on anticipated year of college entry, since Arab women do not serve in the military, and tend to enter college three years younger than Jewish women (as calculated in 1995 census data, and reported by the Israeli Central Bureau of Statistics). Therefore, the first "treated" cohort is the 1971 cohort for Jewish women, and the 1974 cohort for Arab women. Figure (a) shows average college attainment by population group. Figure (b) presents the difference in college attainment between Jewish and Arab women, as well as fitted lines for the pre and post periods. Data from the 2008 Israeli population census.

We further challenge our results by including both the Arab control and the male control in a triple difference specification, shown in Table A4 for college and graduate education. The results hold qualitatively and exhibit very similar magnitudes.

Although we use the Arab control principally to address educational results, Table A3 shows the results for age at first marriage, where we see that Jewish-Israeli women experience a differential increase in age at first marriage, beginning in 1994, compared to Arab-Israeli women.[44] Note, that the estimated effect is considerably larger then the one reported in Table 2, which would be expected since Arab women are not affected in equilibrium by Jewish women postponing marriages, while Jewish men, the marriage partners of Jewish women, may be.

5 Conclusion

Increased access to *in vitro* fertilization offers women the security of a second-line option in case they do not naturally achieve their desired level of fertility. Like any insurance, this guaranteed access to IVF may influence individual behavior: In this case, women delay starting families, using

[44]For spousal income, we did not use men as the principal control group, but rather women who married younger, and thus there is no scope for the alternate Arab control.

Table 10: College Graduation Rates by College Cohort (Arab Control)

	Dependent variable: College Education					
	DiD with GSTT		Slope-Change DiD		GLS Slope-Change DiD	
	(1)	(2)	(3)	(4)	(5)	(6)
jewish × post	0.031	0.030	0.043	0.042	0.038	0.042
	(0.013)**	(0.007)***	(0.012)***	(0.005)***	(0.011)***	(0.010)***
	[0.013]**	[0.013]**	[0.015]***	[0.015]***		
jewish × post × time			-0.005	-0.005	-0.005	-0.005
			(0.002)**	(0.001)***	(0.003)*	(0.002)**
			[0.004]	[0.004]		
jewish × time	0.002	0.002	0.002	0.003	0.003	0.002
	(0.001)***	(0.000)***	(0.001)***	(0.000)***	(0.001)***	(0.001)***
	[0.001]**	[0.001]**	[0.001]**	[0.001]**		
post × time			0.009	0.008	0.010	0.022
			(0.001)***	(0.001)***	(0.002)***	(0.002)***
			[0.002]***	[0.003]***		
post	0.026		0.004		0.003	
	(0.009)***		(0.006)		(0.007)	
	[0.007]***		[0.010]			
jewish	0.243	0.243	0.246	0.246	0.248	0.246
	(0.009)***	(0.005)***	(0.009)***	(0.004)***	(0.008)***	(0.006)***
	[0.027]***	[0.027]***	[0.028]***	[0.028]***		
time	0.004	0.006	0.004	0.004	0.004	0.001
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)**
	[0.000]***	[0.001]***	[0.001]***	[0.001]***		
Constant	0.134	0.171	0.129	0.142	0.129	0.078
	(0.005)***	(0.006)***	(0.005)***	(0.003)***	(0.004)***	(0.006)***
	[0.013]***	[0.018]***	[0.013]***	[0.020]***		
FEs		YES		YES		YES
Observations	125229	125229	125229	125229	54	54
R-Squared	0.0742	0.0748	0.0744	0.0748		

Notes: Columns 1–4: Ordinary least-squares difference-in-differences regression using micro data (no controls included since religiosity and parents' origin controls used only apply to Jewish population). Robust standard errors clustered at the group × cohort level in parentheses; robust standard errors clustered at the group × geography level in square brackets. Columns 5–6: Generalized least squares regression with data collapsed to the population group-college cohort level. Robust standard errors that allow for cross-sectional correlation and for serial correlation with panel specific correlation parameter, in parentheses. Data from the 2008 Israeli population census, restricted to Israeli-born.

*** p<0.01, ** p<0.05, * p<0.1

Table 11: Rates of Graduate Education by Graduate Cohort (Arab Control)

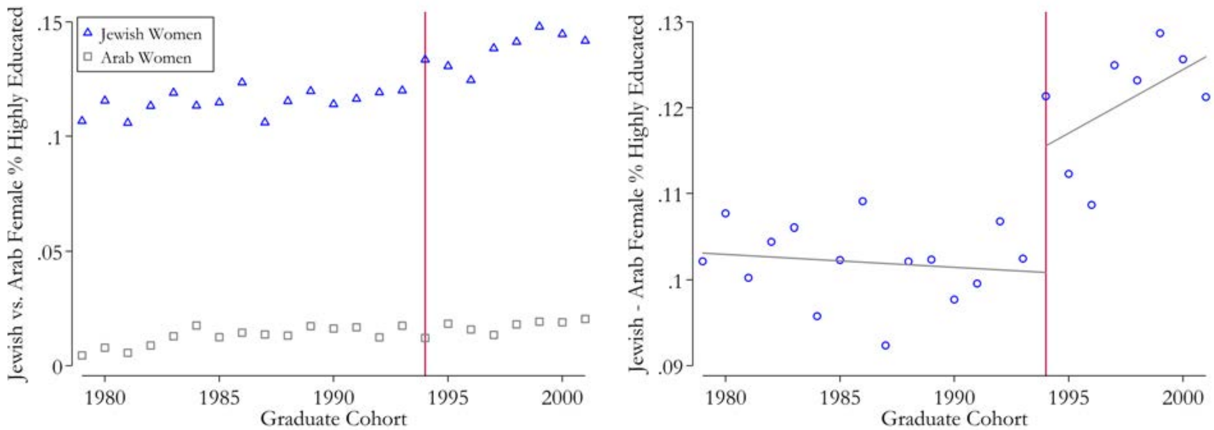
	Dependent variable: Graduate Education					
	DiD with GSTT		Slope-Change DiD		GLS Slope-Change DiD	
	(1)	(2)	(3)	(4)	(5)	(6)
jewish × post	0.022	0.021	0.018	0.018	0.018	0.019
	(0.005)***	(0.003)***	(0.005)***	(0.003)***	(0.004)***	(0.004)***
	[0.007]***	[0.007]***	[0.007]**	[0.007]**		
jewish × post × time			0.002	0.002	0.002	0.002
			(0.001)*	(0.001)**	(0.001)***	(0.001)**
			[0.002]	[0.002]		
jewish × time	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)**	(0.000)	(0.000)
	[0.001]	[0.001]	[0.001]	[0.001]		
post × time			0.000	-0.001	0.000	0.002
			(0.001)	(0.001)	(0.001)	(0.001)***
			[0.001]	[0.001]		
post	-0.005		-0.006		-0.006	
	(0.003)*		(0.003)*		(0.003)*	
	[0.004]		[0.004]			
jewish	0.096	0.096	0.094	0.094	0.094	0.094
	(0.003)***	(0.002)***	(0.003)***	(0.002)***	(0.003)***	(0.003)***
	[0.010]***	[0.010]***	[0.010]***	[0.010]***		
time	0.001	0.001	0.001	0.001	0.001	0.000
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)
	[0.000]***	[0.000]***	[0.000]***	[0.000]**		
Constant	0.022	0.020	0.022	0.020	0.022	0.008
	(0.002)***	(0.001)***	(0.002)***	(0.003)***	(0.002)***	(0.002)***
	[0.003]***	[0.004]***	[0.003]***	[0.005]***		
FEs		YES		YES		YES
Observations	100724	100724	100724	100724	46	46
R-Squared	0.0267	0.0269	0.0267	0.0269		

Notes: Columns 1–4: Ordinary least-squares difference-in-differences regression using micro data (no controls included since religiosity and parents' origin controls used only apply to Jewish population). Robust standard errors clustered at the group × cohort level in parentheses; robust standard errors clustered at the group × geography level in square brackets. Columns 5–6: Generalized least squares regression with data collapsed to the population group-graduate cohort level. Robust standard errors that allow for cross-sectional correlation and for serial correlation with panel specific correlation parameter, in parentheses. Data from the 2008 Israeli population census, restricted to Israeli-born.

*** p<0.01, ** p<0.05, * p<0.1

Figure 18: Percentage of Graduate Education by Graduate Cohort (Arab Control Group)

(a) Separate (b) Difference



Notes: The figure compares graduate education completion rates between Jewish and Arab women over time. The cohorts are aligned based on anticipated year of graduate school entry, since Arab women do not serve in the military, and tend to enter college three years younger than Jewish women (as calculated in 1995 census data, and reported by the Israeli Central Bureau of Statistics). Therefore, the first "treated" cohort is the 1966 cohort for Jewish women, and the 1969 cohort for Arab women. Figure (a) shows average graduate school attainment by population group. Figure (b) presents the difference in graduate school attainment between Jewish and Arab women, as well as fitted lines for the pre and post periods. Data from the 2008 Israeli population census.

the time to pursue additional education and potentially other career opportunities. The delay in starting families is shown by the stark increase in age at first marriage for women following the policy change. The productive use of this time is demonstrated by the rise in completion of college and graduate education.

The policy change may also have impacted men's beliefs about older women's value as partners. We show evidence that older women marry higher quality (richer) partners after the policy. This shift in the marriage equilibrium may further reflect in women's decisions—knowing they will not lose as much reproductive capital by delaying marriage, and that their later-life marriage opportunities will be more favorable as a result, they will have fewer impediments to pursuing desired educational or career investments.

By testing what happens when the threat of later life infertility is attenuated, this research suggests depreciating reproductive capital as a key source of asymmetry between men and women. When better insured against later life infertility, women delay marriage, invest in more education, and marry higher quality partners despite the delay. In the absence of such insurance, this female-specific sharp decline in fertility may contribute to lower human capital investments by women during their reproductive years. In Israel, reproductive capital appears to substantially impact college and graduate education, both because women start families quite young and have relatively high desired fertility rates, and because obligatory military service already delays any decision women make by at least two years. In other OECD countries, however, this investment tradeoff may take place after women have completed their education, when further on-the-job investments are required in order to climb the corporate ladder: late nights at the law firm, medical residencies, or the tenure sprint. Thus, depreciating reproductive capital may help to explain the lack of women in higher-level management positions as well as the high-skill gender wage gap. A wide range of policies, such as increased support for child-rearing in two-career households and access to maternity leave and career re-entry, in addition to access to assisted reproduction technologies, could help alleviate this tradeoff.

In regard to the specific Israeli policy we evaluate, our findings demonstrate that the beneficiary population extends far beyond the women who actually use IVF or other assisted reproduction technologies. Rather, because the guaranteed access acts as insurance in case natural conception fails, all women considering further educational investments or delayed marriage may benefit. This is of critical importance because the cost per user of free IVF with Israel's generous coverage is enormous, and Israel is currently considering measures to limit the policy, having already placed age limits on use, and restricted the number of cycles for certain women. When taking into account the "insurance effect" of the policy, the potential benefits to be weighed against those costs expand considerably.

One slight caution in regards to this cost-benefit calculation is that the type of benefits we describe may not be what the Israeli government had in mind when they enacted the policy. The objectives of the policy were not to increase women's education and career outcomes, but were rather explicitly pro-natalist, aimed at increasing the birth rate of Israeli citizens.[45] Thus, policymakers should note that the behavioral response to IVF access may cause fertility effects to be attenuated, or even go in the opposite direction. If women do delay starting families, assured against the outcome of having zero children, they may nonetheless end up with a smaller overall family size, due to the late start. Moreover, since some evidence suggests individuals are *overly* optimistic about IVF's success rates, some women may delay, fail to conceive naturally, and go on to use the technology, only to be unsuccessful.

These questions of the tradeoff between further human capital investments and labor market productivity versus satisfaction derived from family and home life extend beyond Israeli policy-making. As more and more US companies consider measures such as paying for employees to freeze eggs, which similarly creates insurance against later life infertility, some women who are already planning to delay childbearing may be relieved by the benefit, while others could see a constantly moving finish line for how long they are expected to delay, and feel pressured to submit themselves to intrusive medical procedures and late parenthood. Thus, it is unclear if expanded access to IVF

[45] The policy was defended in courts and described as a part of the fundamental human right to give birth and build a biological family.

is the *best* policy to alleviate the one-sided burden of depreciating reproductive capital. What is clear, however, is that this burden plays a crucial role in women's decisions and outcomes.

References

Abramowitz, J. (2012). Assisted Reproductive Technology and Women's Marriage and Birth Timing: Duration and Competing Risks Analyses. *Working Paper*.

Abramowitz, J. (2014). Turning Back the Ticking Clock: the Effect of Increased Affordability of Assisted Reproductive Technology on Women's Marriage Timing. *Journal of Population Economics 27*(2), 603–633.

Adda, J., C. Dustmann, and K. Stevens (2011). The Career Costs of Children. *CEPR Discussion Paper No. DP8697*.

Andersen, A.-M. N., J. Wohlfahrt, P. Christens, J. Olsen, and M. Melbye (2000, June). Maternal age and fetal loss: population based register linkage study. *BMJ 320*.

Andrews, D. (1993). Test for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica 59*, 817–858.

Avellar, S. and P. J. Smock (2003). Has the Price of Motherhood Declined Over Time? A Cross-Cohort Comparison of the Motherhood Wage Penalty. *Journal of Marriage and Family 65*(3), 597–607.

Bailey, M. J. (2006). More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life Cycle Labor Supply. *Quarterly Journal of Economics 121*(1), 289–320.

Bailey, M. J. (2010). "Momma's Got the Pill": How Anthony Comstock and Griswold v. Conneticut Shaped US Childbearing. *American Economic Review 100*(1), 98–129.

Bernstein, S. (2002). The Rise and Fall of Private Higher Education in Israel. *Policy Studies 52*.

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

Birenbaum-Carmeli, D. (2003). Reproductive Policy in Context: Implications on Women's Rights in Israel, 1945–2000. *Policy Studies 24*(2-3), 101–113.

Birenbaum-Carmeli, D. (2004). 'Cheaper than a Newcomer': on the Social Production of IVF Policy in Israel. *Sociology of Health & Illness 26*(7), 897–924.

Birenbaum-Carmeli, D. (2010). Genetic Relatedness and Family Formation in Israel: Lay Perceptions in the Light of State Policy. *New Genetics and Society 29*(1), 73–85.

Bitler, M. and L. Schmidt (2006). Health Disparities and Infertility: Impacts of State-level Insurance Mandates. *Fertility and Sterility 85*(4), 858–865.

Bitler, M. P. and L. Schmidt (2012). Utilization of Infertility Treatments: the Effects of Insurance Mandates. *Demography 49*(1), 125–149.

Blackburn, M. L., D. E. Bloom, and D. Neumark (1993). Fertility Timing, Wages, and Human Capital. *Journal of Population Economics 6*(1), 1–30.

Bretherick, K. L., N. Fairbrother, L. Avila, S. H. Harbord, and W. P. Robinson (2010). Fertility and Aging: Do Reproductive-aged Canadian Women Know What They Need to Know? *Fertility and Sterility 93*(7), 2162–2168.

Bronson, M. A. and M. Mazzocco (2015). Cohort size and the marriage market: Explaining nearly a century of changes in u.s. marriage rates. *Working Paper*.

Buckles, K. (2007). Stopping the Biological Clock: Infertility Treatments and the Career-family Tradeoff. *Boston University, Unpublished Manuscript*.

Buckles, K. (2008). Understanding the Returns to Delayed Childbearing for Working Women. *The American Economic Review 98*(2), 403–407.

Buckles, K. S. (2013). Infertility Insurance Mandates and Multiple Births. *Health Economics 22*(7), 775–789.

Bundorf, M. K., M. Henne, and L. Baker. Mandated Health Insurance Benefits and the Utilization and Outcomes of Infertility Treatments.

Collins, J. (2001). Cost-effectiveness of in vitro fertilization. In *Seminars in Reproductive Medicine*, Volume 19, pp. 279–289.

Collins, J. A. (2002). An International Survey of the Health Economics of IVF and ICSI. *Human Reproduction Update 8*(3), 265–277.

Danziger, L. and S. Neuman (1999). On the Age at Marriage: Theory and Evidence from Jews and Moslems in Israel. *Journal of Economic Behavior & Organization 40*(2), 179–193.

Dessy, S. and H. Djebbari (2010). High-Powered Careers and Marriage: Can Women Have It All? *The BE Journal of Economic Analysis & Policy 10*(1).

Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2006). Gender Differences in Mate Selection: Evidence From a Speed Dating Experiment. *Quarterly Journal of Economics 121*(2), 637–697.

Goldin, C. and L. F. Katz (2002). The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions. *Journal of Political Economy 110*(4), 730–770.

Gustafsson, S. (2003). Optimal Age at Motherhood. Theoretical and Empirical Considerations on Postponement of Maternity in Europe. In *Family, Household and Work*, pp. 345–367. Springer.

Haimov-Kochman, R., M. Lifshitz, E. Ein Mor, Y. Bdolah, and A. Hurwitz (2012). Misconceptions Regarding Age-Related Fertility Potential Among Israeli Students. *a Research Presented in the Israeli Fertility Association Annual Conference*.

Hamilton, B. H. and B. McManus (2012). The Effects of Insurance Mandates on Choices and Outcomes in Infertility Treatment Markets. *Health Economics 21*(8), 994–1016.

Hashiloni-Dolev, Y. (2013). *The Fertility Revolution*. Ben Shemen: Modan Publishing House, in Hebrew.

Hashiloni-Dolev, Y., A. Kaplan, and S. Shkedi-Rafid (2011). The Fertility Myth: Israeli Students' Knowledge Regarding Age-related Fertility Decline and Late Pregnancies in an Era of Assisted Reproduction Technology. *Human Reproduction 0*(0), 1–9.

Hook, E. B., P. K. Cross, and D. M. Schreinemachers (1983). Chromosomal abnormality rates at amniocentesis and in live-born infants. *JAMA: the journal of the American Medical Association 249*(15), 2034–2038.

Kelsey, T. W. and W. H. B. Wallace (2010, January). Human ovarian reserve from conception to the menopause. *PLoS One 5*.

Kleven, H. J., C. Landais, and J. E. Sogaard (2015). Children and Gender Inequality: Evidence from Denmark. *Working Paper*.

Lahey, J. N. (2012). The Efficiency of a Group-Specific Mandated Benefit Revisited: The Effect of Infertility Mandates. *Journal of Policy Analysis and management 31*(1), 63–92.

Loughran, D. S. and J. M. Zissimopoulos (2009). Why Wait? The Effect of Marriage and Childbearing on the Wages of Men and Women. *Journal of Human Resources 44*(2), 326–349.

Low, C. (2016). Pricing the Biological Clock: Reproductive Capital on the US Marriage Market. *Working Paper*.

Menken, J., J. Trussell, and U. Larsen (1986). Age and infertility. *Science 233*(4771), 1389–1394.

Miller, A. R. (2011). The Effects of Motherhood Timing on Career Path. *Journal of Population Economics 24*(3), 1071–1100.

Mora, J. (1996). The demand for higher education in spain. *European Journal of Education 31*(3), 341–354.

Ohinata, A. (2011). Did the US Infertility Health Insurance Mandates Affect the Timing of First Birth? *CentER Working Paper Series*.

Remennick, L. (2010). Between Reproductive Citizenship and Consumerism: Attitudes towards Assisted Reproductive Technologies among Jewish and Arab Israeli Women. *Kin, Gene, Community: Reproductive Technologies among Jewish Israelis 19*, 318.

Rothman, K. J., L. A. Wise, H. T. Sorensen, A. H. Riis, E. M. Mikkelsen, and E. E. Hatch (2013). Volitional determinants and age-related decline in fecundability: a general population prospective cohort study in denmark. *Fertility and Sterility 99*(7), 1958 – 1964.

Schmidt, L. (2005). Infertility Insurance Mandates and Fertility. *American Economic Review 95*(2), 204–208.

Schmidt, L. (2007). Effects of Infertility Insurance Mandates on Fertility. *Journal of Health Economics 26*(3), 431–446.

Shavit, Y., R. Arum, and A. Gamoran (2007). Stratification in Higher Education: A Comparative Study.

Simonstein, F. (2010). IVF Policies with Emphasis on Israeli Practices. *Health Policy 97*(2), 202–208.

Siow, A. (1998). Differential Fecundity, Markets, and Gender Roles. *Journal of Political Economy 106*(2), 334–354.

Taniguchi, H. (1999). The Timing of Childbearing and Women's Wages. *Journal of Marriage and the Family 61*(4), 1008–1019.

Toner, J. P. (2003, March). Age = egg quality, fsh level = egg quantity. *Fertil Steril 79*.

Velez, M., M. Connolly, I.-J. Kadoch, S. Phillips, and F. Bissonnette (2014). Universal Coverage of IVF Pays Off. *Human Reproduction 29*(6), 1313–1319.

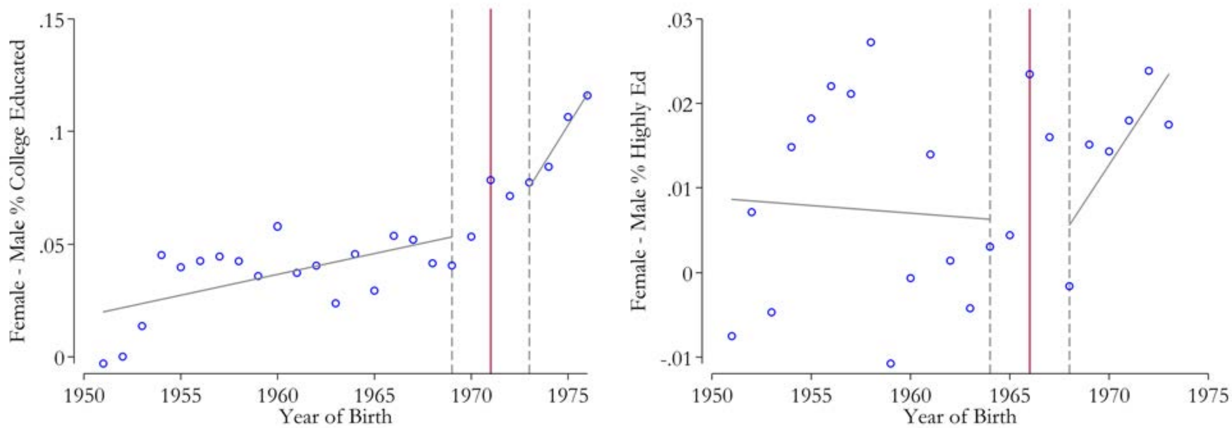
Volanski, A. (2005). *Academy in a Changing Environment*. Kotar, in Hebrew.

Wilde, E. T., L. Batchelder, and D. T. Ellwood (2010). The Mommy Track Divides: The Impact of Childbearing on Wages of Women of Differing Skill Levels. *Working Paper* (16582).

6 Appendix

Figure 19: Difference in Education Rates with Fuzzy Treatment

(a) College Education (b) Graduate Education



Notes: These graphs show that the difference in education outcomes post-policy is still apparent with a two-year interval on either side of our “most treated” cohort. Older cohorts may have been partially treated, if some individuals had not yet made their educational decisions and thus were able to respond to the policy change, and younger cohorts may not have been fully treated, if some individuals had made their educational decisions already. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

Table A1: College and Graduate Graduation Rates, Men's Cohort Adjusted

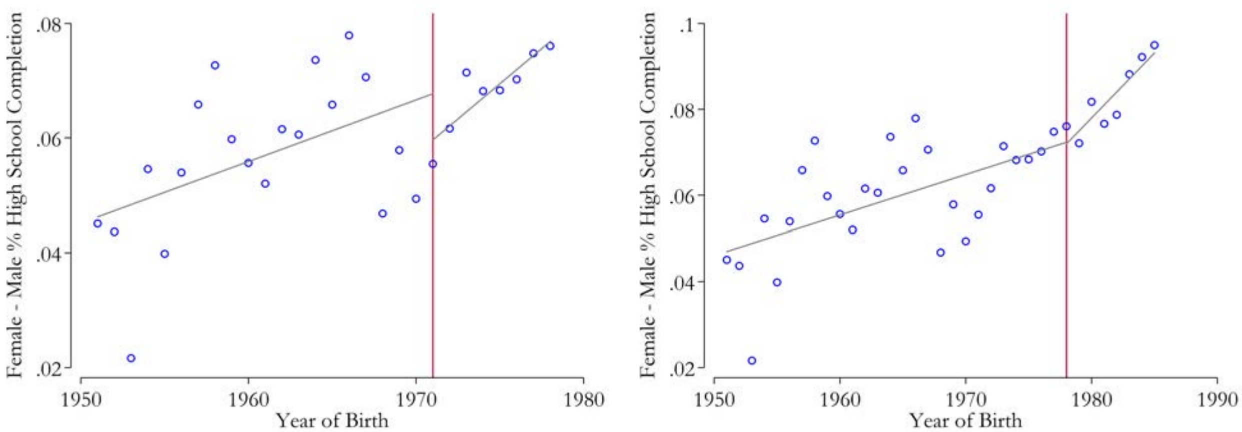
Dependent variable:	College Education		Graduate Education	
	DiD with GSTT			
	(1)	(2)	(3)	(4)
fem × post	0.024	0.018	0.014	0.009
	(0.007)***	(0.006)***	(0.006)**	(0.004)**
	[0.016]	[0.015]		
fem × time	0.002	0.003	-0.000	-0.000
	(0.000)***	(0.000)***	(0.000)	(0.000)
	[0.001]**	[0.001]***		
post	-0.005		0.010	
	(0.004)		(0.004)**	
	[0.012]			
female	0.060	0.063	0.003	0.005
	(0.004)***	(0.004)***	(0.004)	(0.003)*
	[0.025]**	[0.025]**		
time	0.004	0.005	0.000	0.001
	(0.000)***	(0.001)***	(0.000)	(0.000)***
	[0.001]***	[0.001]***		
Constant	0.497	0.495	0.195	0.200
	(0.007)***	(0.008)***	(0.005)***	(0.004)***
	[0.031]***	[0.031]***		
FEs (YOB)	YES		YES	
FEs (Cohort)		YES		YES
Observations	171617	171617	136998	136998
R-Squared	0.110	0.110	0.0468	0.0469

Notes: Columns 1–4: Ordinary least-squares difference-in-differences regression using micro data, including controls for religiosity and parents' origin. Robust standard errors clustered at the gender × cohort level in parentheses; robust standard errors clustered at the gender × geography level in square brackets. Columns 5–6: Generalized least squares regression with data collapsed to the gender-education cohort level. Robust standard errors that allow for cross-sectional correlation and for serial correlation with panel specific correlation parameter, in parentheses. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

*** p<0.01, ** p<0.05, * p<0.1

Figure A1: High School Placebo

(a) 1971 Cutoff (b) 1978 Cutoff



Notes: We test for a change in high school completion rates to rule out broader increases in education driving our effects. We test this for both the 1971 cohort, which would have been affected if the impact was something that affected all individuals born in 1971, and the 1978 cohort, which would have been affected if the impact affected all individuals pursuing schooling in 1994. Figure (a) shows average high school completion for women and men by birth cohort. Figure (b) presents the difference in high school completion between women and men, as well as fitted lines for the pre and post periods. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

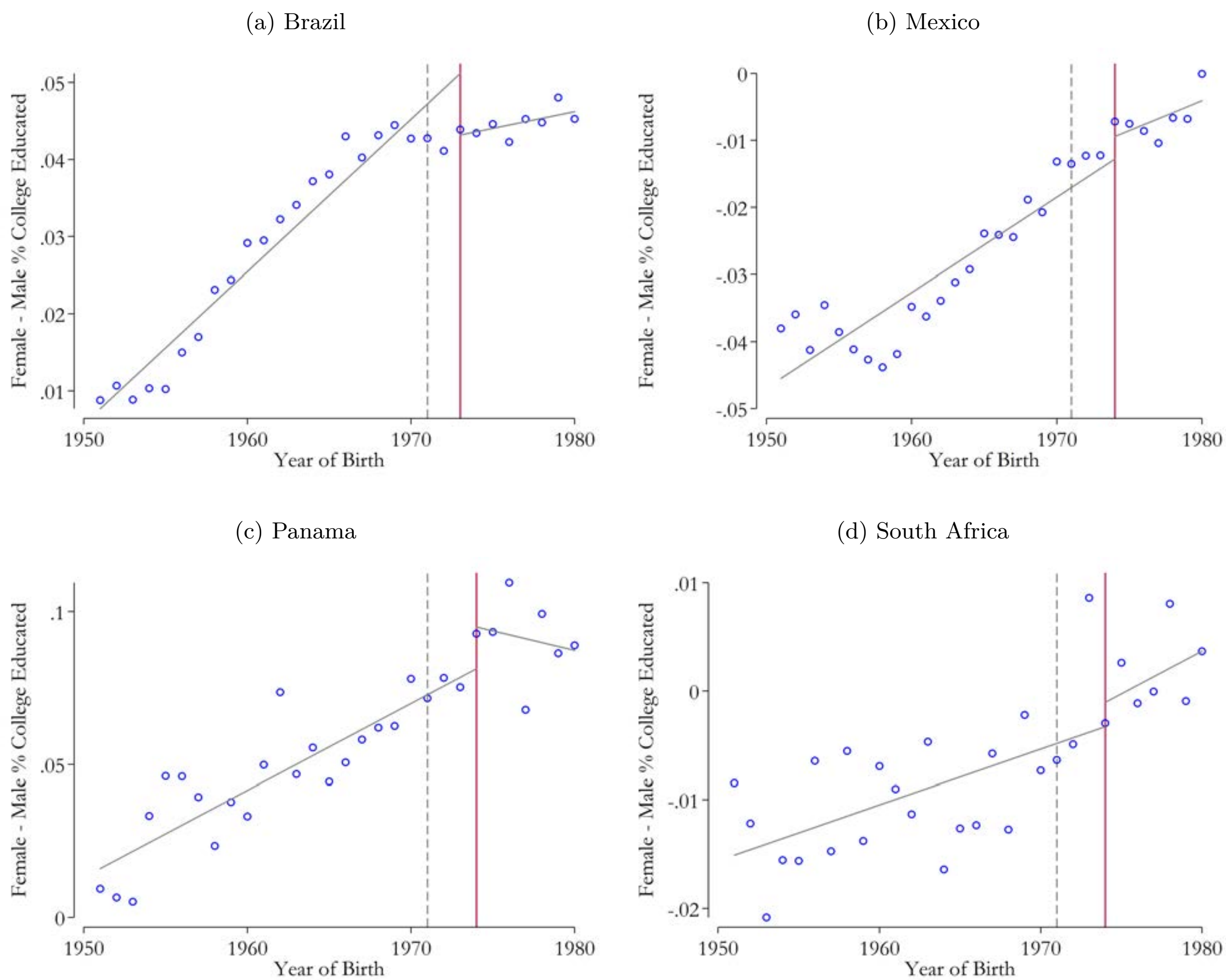
Table A2: Alternative Spousal Quality Specifications, Marrying Older vs. Younger

Dep. Variable:	Income Age Residual		Income		College Education	
	With Wife Controls		With Spouse YOB FEs			
	(1)	(2)	(3)	(4)	(5)	(6)
older × post	25192.8	20460.6	23717.2	26113.8	0.042	0.043
	(7943.6)***	(7011.0)***	(7688.3)***	(6845.9)***	(0.020)**	(0.017)**
	[7929.0]***	[8234.9]**	[7958.3]***	[8315.6]***	[0.019]**	[0.019]**
married older	-28149.4	-25708.8	-24185.2	-25643.4	-0.054	-0.054
	(9113.4)***	(6408.1)***	(5899.1)***	(5807.1)***	(0.014)***	(0.013)***
	[15926.7]*	[10611.3]**	[13270.7]*	[12930.7]**	[0.028]*	[0.028]*
post	5627.8		1478.5		0.034	
	(6397.4)		(6349.5)		(0.020)*	
	[5993.9]		[6370.5]		[0.018]*	
time	-2236.9	-1338.3	-2646.8	-1849.5	0.001	0.002
	(1416.7)	(508.8)**	(767.0)***	(573.4)***	(0.001)	(0.001)***
	[1569.3]	[738.0]*	[660.6]***	[786.7]**	[0.001]	[0.002]
Constant	26969.5	18217.1	80667.3	68557.7	0.553	0.568
	(50020.1)	(6181.8)***	(6163.6)***	(6816.0)***	(0.016)***	(0.013)***
	[63463.1]	[12333.9]	[7553.9]***	[9372.6]***	[0.024]***	[0.029]***
FEs (YOM)		YES		YES		YES
Observations	19458	17474	19458	19458	20542	20542
R-Squared	0.0945	0.0968	0.0556	0.0568	0.0795	0.0805

Notes: Ordinary least-squares difference-in-differences regression using micro data, including controls for religiosity and parents' origin. Robust standard errors clustered at the age group × year level in parentheses; robust standard errors clustered at the age group × geography level in square brackets. Column 1 controls for wife's age and education, Column 2 controls for education and income in addition to year-of-marriage fixed effects. Columns 3–4 use raw income as the dependent variable, controlling for spouse's age through year-of-birth fixed effects. Columns 5–6 use the same specification as in table 8 columns 1–2 with an indicator for spousal college education as outcome. Data from the 2008 Israeli population census, restricted to Israeli-born Jews.

*** p<0.01, ** p<0.05, * p<0.1

Figure A2: College Attainment by Birth Cohort in Comparable Countries

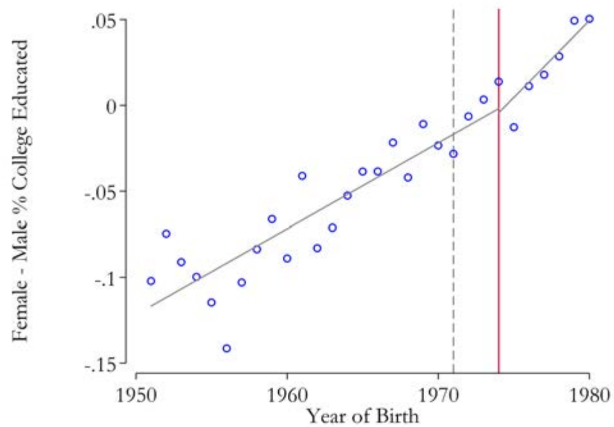


Notes: We use countries that have censuses around the time of Israel's 2008 Census and similar GDP per capita as Israel to conduct placebo tests, showing that the cohort entering college in 1994 in other countries was not similarly affected (in each country, the red line is shifted according to typical college entry age of students in that country). This would be the case if broader international shifts in the nineties were responsible for the effects that we see. We do not observe similar discontinuous increases in female versus male college attainment over time in any of the other countries. The same lack of discontinuous trends is true when looking at graduate school attainment as well, although the data is somewhat noisier. Data obtained from IPUMS. Census data for Brazil, Mexico, and Panama are from 2010 and census data for South Africa is from 2007.

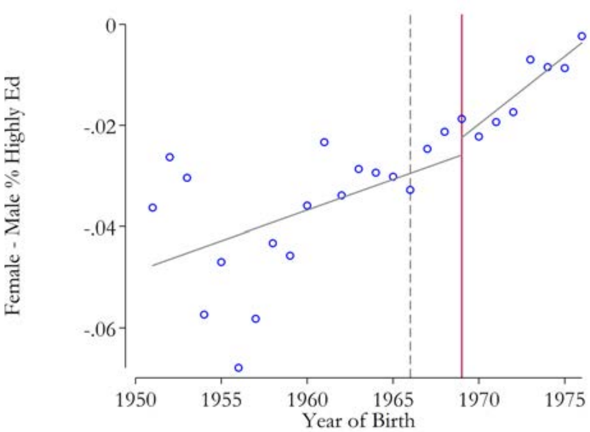
Figure A3: Placebo Test using Arab-Israelis

(a) Age First Marriage

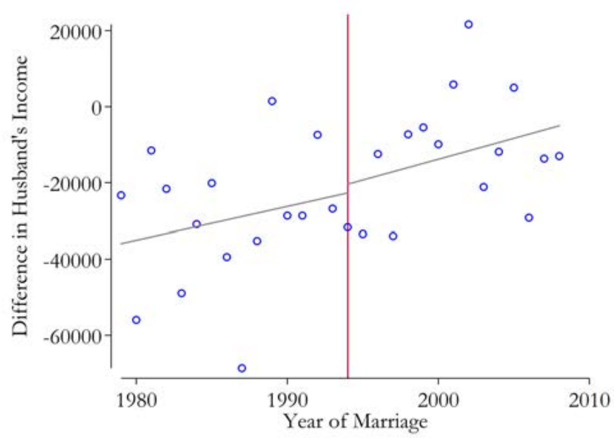
(b) College Education



(c) Graduate Education



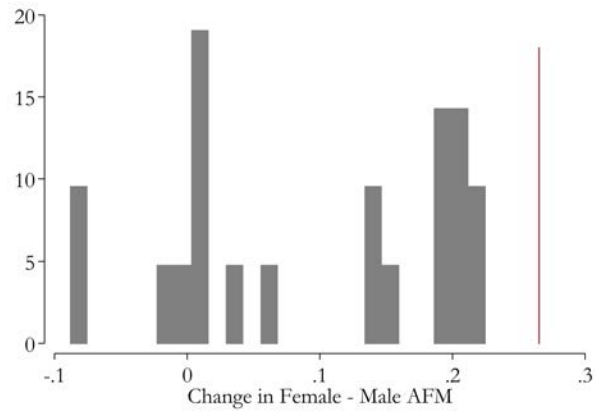
(d) Husband's Income



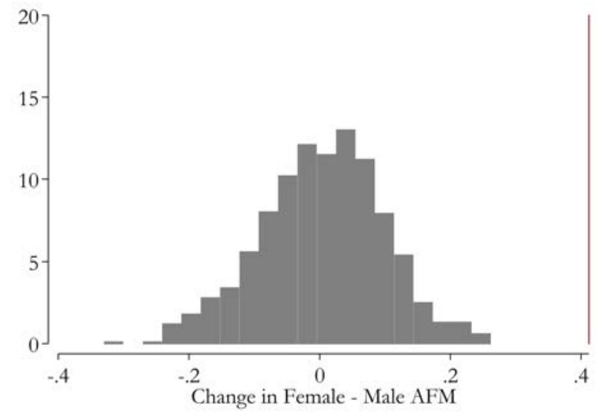
Notes: The figure shows our four key comparisons using Arab-Israelis, who are less likely to be impacted by the policy change due to religious restrictions on IVF use and lower baseline educational attainment, as a placebo test for non-IVF related changes in Israel that may have affected all women. Figure (a) shows the difference in men and women's age at first marriage by marriage year, figures (b) and (c) show college and graduate school attainment, respectively, by birth year, with a red line at the cohort expected to be entering in 1994, and a dashed line at the cohort considered "treated" for Jewish individuals (due to a three-year difference in age at college entry between Jewish and Arab Israelis), and figure (d) shows the difference in husband's income-age-residual for women who marry at or above 30 versus women who marry younger. Data from 2008 Israeli population census.

Figure A4: Permutations for Age at First Marriage

(a) 5-Year Treatment Period



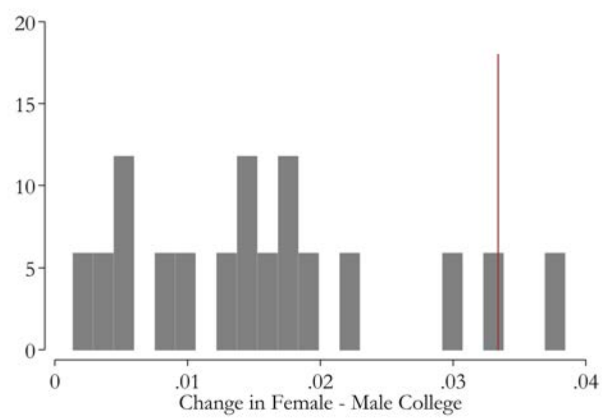
(b) 1000 Permutatins with Random Treatment



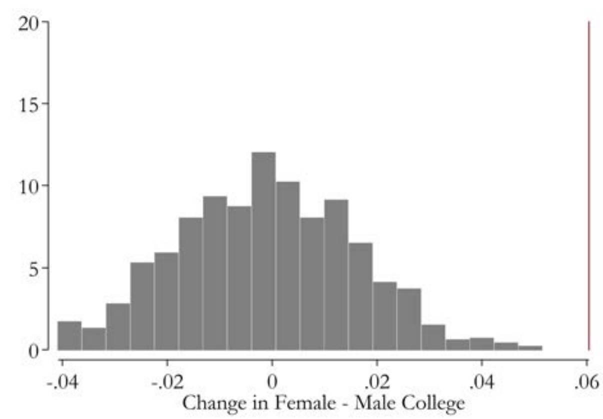
Notes: The figure on the left is created by running a similar regression as our column 1 specification, except with a ten year data period, with five years control and five years treatment, sequentially, for every possible ten year period in our data range. The red line represents the effect size of the actual treatment year, with this ten-year data period (the ten-year approach allows us to compare our actual treatment to other break points, with the same number of years before and after). The figure at right uses the same number of "treated" years as in the true model, but randomly draws them from the study period (for an example of this approach, see Agarwal et. al, 2015). We perform 1,000 such random draws.

Figure A5: Permutations for College Education

(a) 5-Year Treatment Period



(b) 1000 Permutatins with Random Treatment

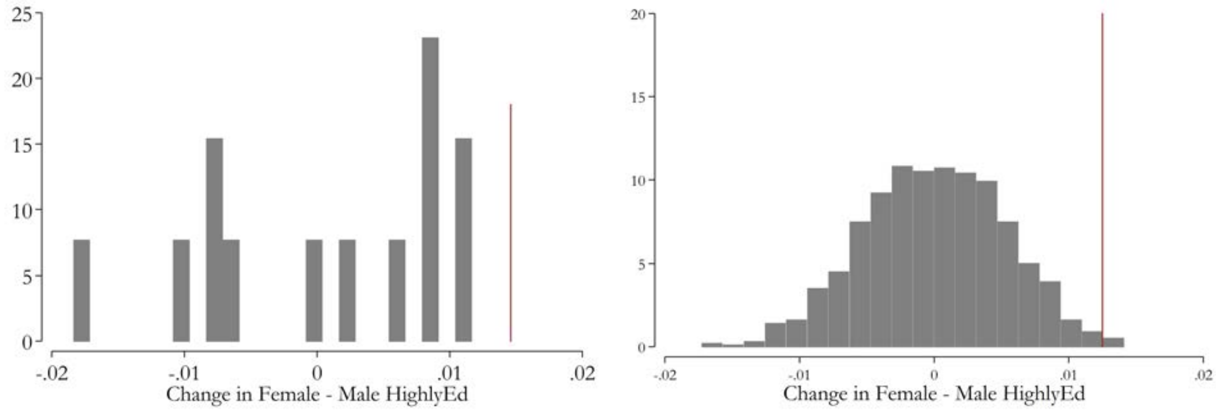


Notes: The figure on the left is created by running a similar regression as our column 1 specification, except with a ten year data period, with five years control and five years treatment, sequentially, for every possible ten year period in our data range. The red line represents the effect size of the actual treatment year, with this ten-year data period (the ten-year approach allows us to compare our actual treatment to other break points, with the same number of years before and after). The figure at right uses the same number of "treated" years as in the true model, but randomly draws them from the study period (for an example of this approach, see Agarwal et. al, 2015). We perform 1,000 such random draws.

Figure A6: Permutations for Graduate Education

(a) 5-Year Treatment Period

(b) 1000 Permutatins with Random Treatment

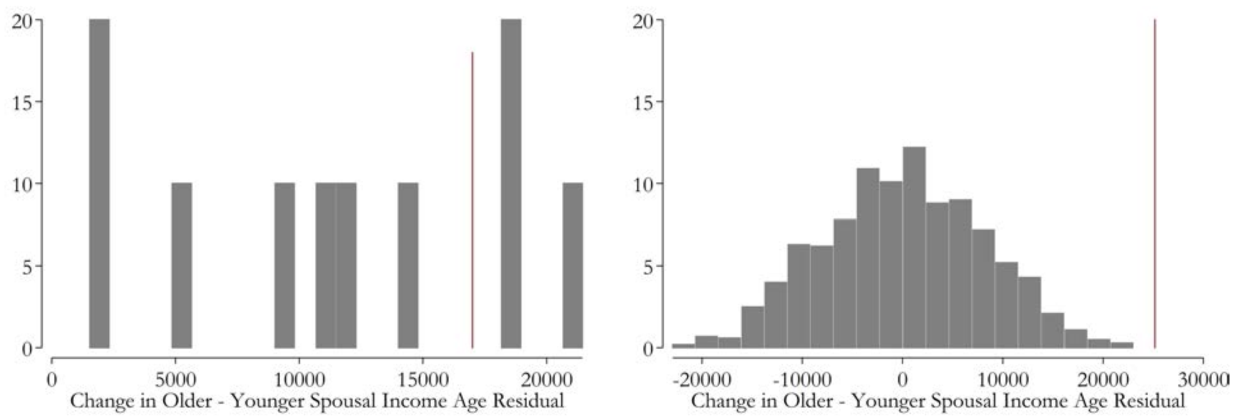


Notes: The figure on the left is created by running a similar regression as our column 1 specification, except with a ten year data period, with five years control and five years treatment, sequentially, for every possible ten year period in our data range. The red line represents the effect size of the actual treatment year, with this ten-year data period (the ten-year approach allows us to compare our actual treatment to other break points, with the same number of years before and after). The figure at right uses the same number of "treated" years as in the true model, but randomly draws them from the study period (for an example of this approach, see Agarwal et. al, 2015). We perform 1,000 such random draws.

Figure A7: Permutations for Spousal Income Age Residual

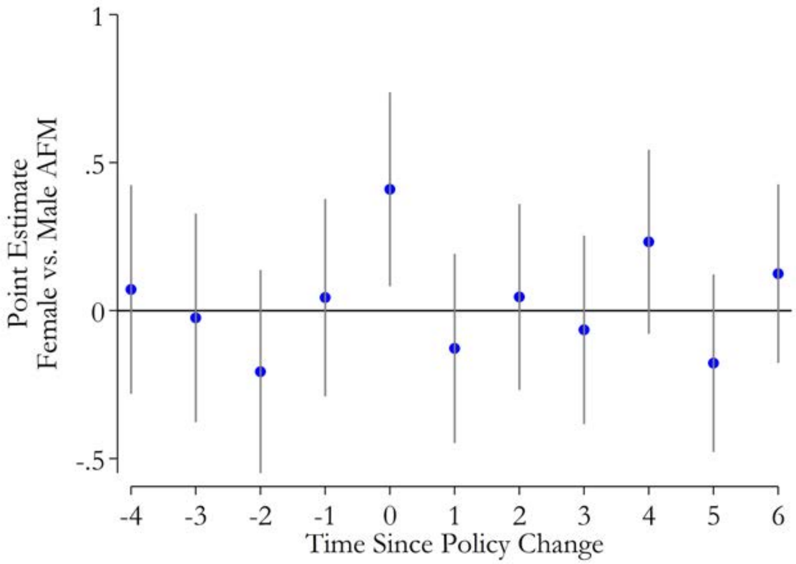
(a) 5-Year Treatment Period

(b) 1000 Permutatins with Random Treatment



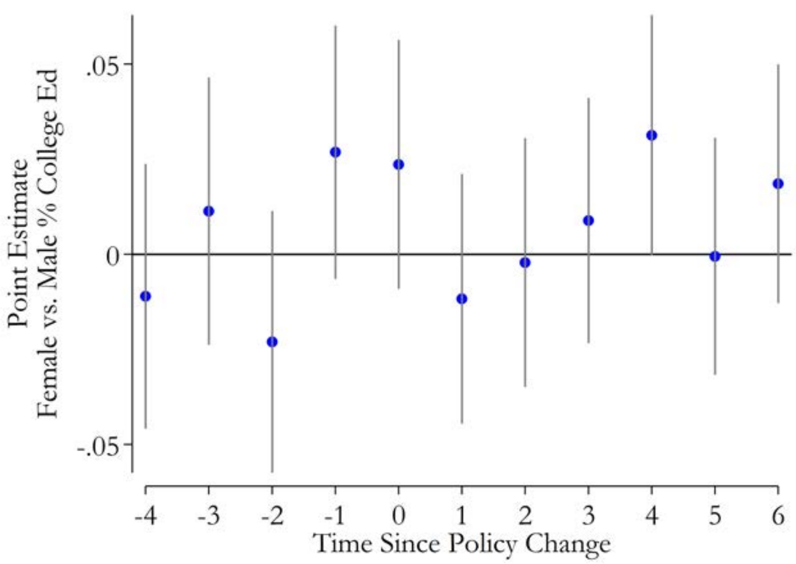
Notes: The figure on the left is created by running a similar regression as our column 1 specification, except with a ten year data period, with five years control and five years treatment, sequentially, for every possible ten year period in our data range. The red line represents the effect size of the actual treatment year, with this ten-year data period (the ten-year approach allows us to compare our actual treatment to other break points, with the same number of years before and after). The figure at right uses the same number of "treated" years as in the true model, but randomly draws them from the study period (for an example of this approach, see Agarwal et. al, 2015). We perform 1,000 such random draws.

Figure A8: Event Study: Age of First Marriage (Distributed lags)



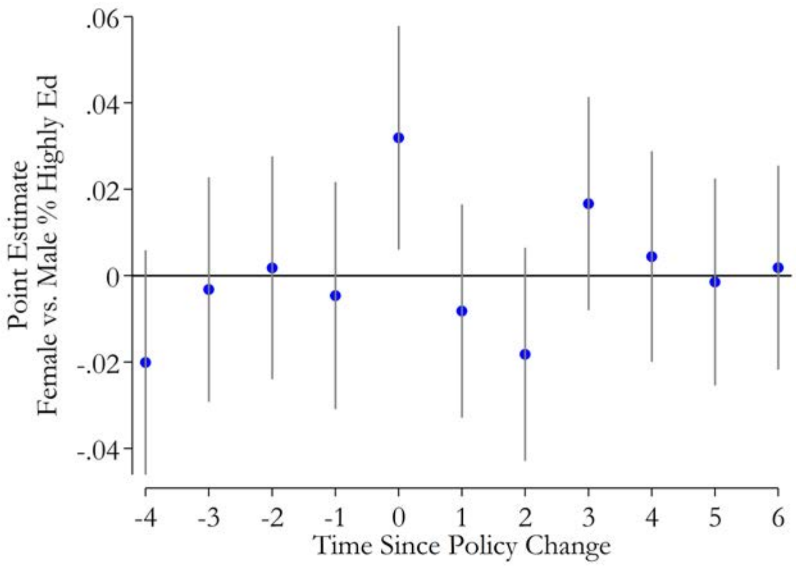
Notes: The figure presents point estimates and confidence intervals for coefficients on a series of dummy variables for being married in or after the specified year interacted with a dummy variable for female. The outcome variable is age at first marriage. 1994 is the time of the policy change, noted as time 0 on the x-axis.

Figure A9: Event Study: College Education (Distributed lags)



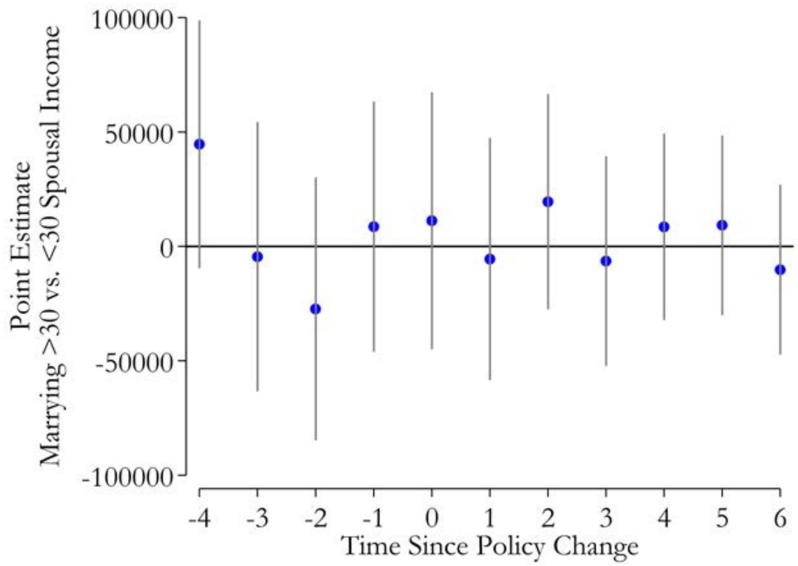
Notes: The figure presents point estimates and confidence intervals for coefficients on a series of dummy variables for having a birth year at or above the specified year (i.e., being in or younger than the specified cohort) interacted with a dummy variable for female. The outcome variable is an indicator for college graduation. The 1971 cohort represents the first affected cohort and hence is the “time” of the policy change, noted as time 0 on the x-axis.

Figure A10: Event Study: Graduate Education (Distributed lags)



Notes: The figure presents point estimates and confidence intervals for coefficients on a series of dummy variables for having a birth year at or above the specified year (i.e., being in or younger than the specified cohort) interacted with a dummy variable for female. The outcome variable is an indicator for post-college graduation. The 1966 cohort represents the first affected cohort and hence is the "time" of the policy change, noted as time 0 on the x-axis.

Figure A11: Event Study: Spousal Income Age Residual, 1984-2003 (Distributed lags)



Notes: The figure presents point estimates and confidence intervals for coefficients on a series of dummy variables for being married in or after the specified year interacted with a dummy variable for being 30 or older at the year of marriage. The outcome variable is spousal income. 1994 is the time of the policy change, noted as time 0 on the x-axis. The sample is restricted to women with spousal matches that got married when they were older than 24 but younger than 35.

Table A3: Age at First Marriage (Arab Control)

	Dependent variable: Age First Marriage					
	DiD		Slope-Change DiD		GLS Slope-Change DiD	
	(1)	(2)	(3)	(4)	(5)	(6)
jewish × post	1.019	1.016	-0.076	-0.057	-0.033	0.488
	(0.127)***	(0.106)***	(0.152)	(0.082)	(0.150)	(0.325)
	[0.208]***	[0.208]***	[0.192]	[0.192]		
jewish × post × time			0.058	0.059	0.056	0.349
			(0.016)***	(0.010)***	(0.020)***	(0.062)***
			[0.030]*	[0.030]*		
jewish × time			0.047	0.045	0.047	-0.114
			(0.009)***	(0.004)***	(0.013)***	(0.042)***
			[0.018]**	[0.018]**		
post × time			-0.020	-0.029	-0.022	-0.304
			(0.013)	(0.007)***	(0.021)	(0.048)***
			[0.022]	[0.027]		
post	-0.870		-0.098		-0.101	
	(0.162)***		(0.122)		(0.155)	
	[0.198]***		[0.169]			
jewish	2.066	2.064	2.374	2.365	2.358	0.698
	(0.074)***	(0.050)***	(0.066)***	(0.036)***	(0.111)***	(0.257)***
	[0.283]***	[0.280]***	[0.386]***	[0.384]***		
time	0.125	0.101	0.080	0.086	0.080	0.193
	(0.007)***	(0.009)***	(0.006)***	(0.004)***	(0.014)***	(0.032)***
	[0.014]***	[0.010]***	[0.016]***	[0.018]***		
Constant	21.984	21.564	21.688	21.626	21.692	23.260
	(0.084)***	(0.127)***	(0.036)***	(0.033)***	(0.116)***	(0.011)***
	[0.269]***	[0.243]***	[0.334]***	[0.355]***		
FEs		YES		YES		YES
Observations	124744	124744	124744	124744	62	62
R-Squared	0.121	0.122	0.122	0.123		

Notes: Columns 1–4: Ordinary least-squares difference-in-differences regression using micro data (no controls included since religiosity and parents' origin controls used only apply to Jewish population). Robust standard errors clustered at the group × year level in parentheses; robust standard errors clustered at the group × geography level in square brackets. Columns 5–6: Generalized least squares regression with data collapsed to the group-year level. Robust standard errors that allow for cross-sectional correlation and for serial correlation with panel specific correlation parameter, in parentheses. Data from the 2008 Israeli population census, restricted to Israeli-born.

*** p<0.01, ** p<0.05, * p<0.1

Table A4: College and Graduate Education Rates (Arab Control), Triple Difference

Dependent variable:	College Education		Graduate Education	
	DiD with GSTT		DiD with GSTT	
	(1)	(2)	(3)	(4)
jewish × post × fem	0.028	0.027	0.010	0.010
	(0.012)**	(0.012)**	(0.009)	(0.009)
	[0.013]**	[0.013]**	[0.010]	[0.010]
jewish × time × fem	-0.004	-0.004	-0.002	-0.003
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
	[0.001]***	[0.001]***	[0.001]***	[0.001]***
jewish × fem	0.051	0.051	0.028	0.028
	(0.008)***	(0.008)***	(0.005)***	(0.005)***
	[0.012]***	[0.012]***	[0.007]***	[0.007]***
jewish × time	0.006	0.006	0.002	0.003
	(0.001)***	(0.001)***	(0.001)***	(0.000)***
	[0.001]***	[0.001]***	[0.001]***	[0.001]***
jewish × post	0.003	0.002	0.011	0.011
	(0.015)	(0.011)	(0.007)	(0.006)*
	[0.015]	[0.015]	[0.008]	[0.008]
fem × time	0.007	0.007	0.002	0.002
	(0.001)***	(0.001)***	(0.000)***	(0.000)***
	[0.001]***	[0.001]***	[0.001]***	[0.001]***
fem × post	0.002	0.001	0.004	0.004
	(0.009)	(0.009)	(0.005)	(0.005)
	[0.010]	[0.010]	[0.006]	[0.006]
post	0.024		-0.010	
	(0.010)**		(0.005)**	
	[0.007]***		[0.004]**	
female	0.005	0.005	-0.026	-0.026
	(0.006)	(0.006)	(0.003)***	(0.003)***
	[0.010]	[0.010]	[0.004]***	[0.004]***
jewish	0.192	0.192	0.069	0.069
	(0.011)***	(0.008)***	(0.005)***	(0.003)***
	[0.026]***	[0.026]***	[0.014]***	[0.014]***
time	-0.002	-0.002	-0.001	-0.002
	(0.001)***	(0.001)*	(0.000)***	(0.000)***
	[0.001]***	[0.001]**	[0.000]**	[0.000]***
Constant	0.129	0.155	0.048	0.044
	(0.007)***	(0.008)***	(0.003)***	(0.001)***
	[0.006]***	[0.008]***	[0.004]***	[0.005]***
FEs		YES		YES
Observations	248081	248081	199129	199129
R-Squared	0.0554	0.0558	0.0191	0.0192

Notes: Columns 1–4: Ordinary least-squares difference-in-differences regression using micro data (no controls included since religiosity and parents' origin controls used only apply to Jewish population). Robust standard errors clustered at the group × year level in parentheses; robust standard errors clustered at the group × geography level in square brackets. Columns 5–6: Generalized least squares regression with data collapsed to the gender-population group-college cohort level. Robust standard errors that allow for cross-sectional correlation and for serial correlation with panel specific correlation parameter, in parentheses. Data from the 2008 Israeli population census, restricted to Israeli-born.

*** p<0.01, ** p<0.05, * p<0.1