Can Women Have Children and a Career? IV Evidence from IVF Treatments*

Petter Lundborg, Erik Plug and Astrid Würtz Rasmussen

October 2016

Abstract

This paper introduces a new IV strategy based on IVF (in vitro fertilization) induced fertility variation among childless women to estimate the causal effect of having children on their career. For this purpose, we use administrative data on IVF treated women in Denmark. Because observed chances of IVF success do not depend on labor market histories, IVF treatment success provides a plausible instrument for childbearing. Our IV estimates indicate that fertility effects on earnings are: (a) negative, large and long lasting; (b) driven by fertility effects on hourly earnings and not so much on labor supply; and (c) much stronger at the extensive margin than at the intensive margin.

In almost all labor markets, women with children work and earn less than women without children (Martin Browning 1992; Claudia Goldin 1992; Jane Waldfogel 1998; James Feyrer, Bruce Sacerdote and Ariel Stern 2008; Marianne Bertrand 2011). There are two leading explanations for these labor market differences. The first one is based on causation; that is, having children has adverse labor market consequences for women. The second one is based on adverse selection; that is, women with children work and earn less, regardless of having children.

Separating causation from adverse selection is of interest to researchers, policy makers, and parents. Researchers know little, if anything, about the causal labor market consequences of having children. Identifying the labor market effects of having children (the extensive fertility margin), as opposed to the labor market effects of having additional children among women who already have children (the intensive fertility margin), has proved very difficult. Policy makers often base their family-friendly policies (intended to support the labor supply of women with children) on beliefs that children seriously hinder the labor market career of women. These policies may be less efficient if women with children are also women with weaker labor market potential. Similarly, parents may want to know when they decide to have children. After all, the labor market response to having children captures a response to which all parents are (or have ever been) exposed.

In this paper we examine the labor market response to having children among women who have no children (yet). To identify fertility effects measured at the extensive margin, we propose a novel instrumental variable strategy based on in vitro fertilization (IVF). In particular, we sample childless women who go through IVF, treat IVF treatment success at the first IVF treatment as a natural experiment, and subsequently compare the working careers of women after a successful first treatment (treatment group) to those of women after a failed first treatment (comparison group). Since observed working histories of successfully and unsuccessfully treated women are virtually identical before they seek IVF treatment, we believe that IVF treatment success creates exogenous variation in the likelihood of having children and can be used as an instrumental variable to estimate the causal effect of having any children on a wide range of female labor market outcomes.

The data we use are drawn from multiple administrative registers in Denmark. IVF information comes from the IVF registers which cover information on all fertility treatments (with fertility outcomes) taking place in public and

private fertility clinics in Denmark between the years 1994 and 2005. Labor market information comes from tax and employer-employee registers which hold records on annual earnings (including parental leave and sickness benefits), annual labor supply, time out of work, hourly earnings as well as job holdings for a 19 year period running from 1991 to 2009. At Statistics Denmark these registers are further matched and complemented with standard demographic variables, such as education, age, gender, marital status, and total number of children.

Our main finding is that women who are successfully treated by IVF earn persistently less because of having children. We explain the decline in annual earnings by women working less when children are young and getting paid less when children are older. We explain the decline in hourly earnings, which is often referred to as the motherhood penalty, by women moving to lower-paid jobs that are closer to home. Equally interesting is our finding that the decline in annual earnings we estimate for having children is far greater than the decline in earnings we get when we estimate the effect of having additional children among women who already have children. To put our findings into context, we also estimate the effect of having more children on earnings using the more commonly used instrumental variable strategy based on twin births. We find estimates that are generally smaller than those from the United States, which has, by Danish standards, meager maternity leave and child care benefits.

The remainder of the paper proceeds as follows. Section I provides the literature background and motivation behind this study. Section II discusses context and data. Section III introduces our empirical strategy. Section IV reports on the labor market consequences of IVF treatments (in terms of reduced-form results and instrument validity tests). Section V presents our main set of results. Section VI compares fertility effects measured at the extensive and intensive margins. Section VII concludes.

I Literature

To estimate the labor market consequences of childbearing, empirical economists have turned to natural experiments and looked for variables that induce variation in the number of children for reasons unrelated to labor market outcomes. Two such variables, which are arguably unrelated to the preferences and abilities of parents, are often used to estimate the effect of children on

various labor supply choices (including participation, hours worked and annual labor earnings): twins at first birth and the sex composition of the first two children (in combination with parental preferences for mixed sex siblings). Those empirical studies that rely on the twin experiment typically find that mothers with twins work less than mothers with singletons when twins are young, but the same mothers work as much, if not somewhat more, when their twins get older (Mark Rosenzweig and Kenneth Wolpin 1980; Stephen Bronars and Jeff Grogger 1994; Joyce Jacobsen, James Pearce and Julian Rosenbloom 1999; and James Vere 2011). Those empirical studies that rely on the same sex siblings experiment find that mothers of same sex children work less than mothers of mixed sex children because they are more likely to have a third child. The same mothers, however, seem to catch up in the longer run (Joshua Angrist and William Evans 1998; Maria Iacovou 2001; Guillermo Cruces and Sebastian Galiani 2007; Eric Maurin and Julie Moschion 2009; Lalaina Hirvonen 2009; Nikolay Angelov and Arizo Karimi 2012). On the whole, these studies suggest that mothers work less because of childbearing, but that the observed fertility effects are relatively small and mostly short-lived.

Although the twin and same sex experiments provide valuable and credible effect estimates, their informational value is limited to fertility effects at the intensive margin. These experiments create exogenous variation in the number of children, but only among women who already have children; that is, twins at first birth raise the likelihood of going from one to two or more children, whereas same sex children raise the likelihood of going from two to three or more children. The question we ask here is how fertility affects the labor market career at the extensive margin, where childless women become mothers or not. While standard theories of household production, child quality and economies of scale predict that labor market consequences are stronger at the extensive fertility margin, empirical evidence is rare.

There have been a few attempts to analyze the causal relationship between fertility at the extensive margin and female labor supply. The recent studies on this question, that exploit natural experiments, have been limited in their causal research design or data availability (Jorge Agüero and Mindy Marks

4

2011; Julian Cristia 2008).[1,2]

The current literature on the extensive fertility margin is in its infancy with many unresolved issues. In this paper, it is our goal to explore some of these unresolved but promising issues; that is, we apply a novel IV strategy using a very large sample of IVF treated families to measure the short, medium, and long run consequences of childbearing at both extensive and intensive margins.

II Institutions and Data

In this section we first describe Danish family-friendly policies (partly intended to support female labor supply) to provide the institutional context in which we estimate the causal effect of having children on female labor supply and earnings. We then show how IVF data can help us to estimate these fertility effects by presenting more details on the IVF register, IVF treatments, the IVF instrument, and the primary IVF sample we use in our analysis.

A Family-Friendly Policies

Denmark is known for its family-friendly policies (including paid maternity and parental leave, job-protection and subsidized child care arrangements). Over the period we consider, parents are covered first by paid and job-protected leave and then by public child care. With regard to paid and job-protected leave policies, parents are entitled to 22 weeks of maternity leave (including 4 weeks of pregnancy leave), 2 weeks of paternity leave, and 32 weeks of shared

---

[1] Agüero and Marks (2011) treat self-assessed infertility as a natural experiment and compare fertility rates and labor supply responses between infertile and fertile women aged 20 to 44 in 26 developing countries. While the authors find that fertile women work as much as infertile women, regardless of the fertility margin, they also find that infertility is more common among better educated women who live in cities. If infertility in women is not randomly assigned, it is not clear whether self-assessed infertility is the most appropriate instrument to use in an instrumental variable strategy. Cristia (2008) focuses on childless women seeking help to achieve pregnancy in the United States. In his context, fertility treatments include simple and inexpensive procedures such as medical advice and fertility tests, but exclude IFV treatments, which play a key role in our study. With treatment success taken as a natural experiment, he finds that women with infants work significantly less than those without. Because of data limitations, however, Cristia can only consider labor supply responses of mothers with children younger than one year.

[2] Parallel to the few studies based on natural experiments, two recent studies take a structural or event study approach to estimate the labor market consequences of parenthood. Jérôme Adda, Christian Dustmann, and Katrien Stevens (forthcoming) use a dynamic life-cycle career model allowing for endogenous fertility. Henrik Kleven, Camille Landais, and Jakob Søgaard (2016) use an event study model around the birth of the first child assuming exogenous timing of childbearing.

parental leave. Before 2002, these paid leave arrangements were less generous and restricted to 18 weeks of maternity leave (including 4 weeks of pregnancy leave), 2 weeks of paternity leave, and 10 weeks of shared parental leave (Helena Skyt Nielsen, Marianne Simonsen and Mette Verner 2004). Parental leave payments depend on previous earnings and differ by sector of employment. In the public sector, the compensation rate is a 100 percent (for the first 24 weeks). In the private sector, the compensation rate can vary by workplace and averages out at 60 to 70 percent (Elina Pylkkänen and Nina Smith 2004). With regard to child care policies, parents are entitled to subsidized public child care when children are somewhere between 6 and 12 months old (Simonsen 2010).

B   IVF Register

The IVF register, held by the Danish National Board of Health, collects information on all IVF treatments taking place in public and private fertility clinics and hospitals. Between the years 1994 and 2005, reporting to the IVF register was mandatory. Thus, this register captures close to 100 percent of IVF treatments. The IVF register includes information on the reason for infertility, the mode of treatment, the number of eggs retrieved from the womb, the number of fertilized eggs transferred back, treatment outcome (birth, abortion, stillbirth, or failure), the date of treatment and where applicable the date of birth. The IVF register has been merged to other administrative registers to get longitudinal information running from 1991 to 2009 on standard demographic variables, including education, age, gender, marital status, number of children and labor market variables, including labor force status, the number of hours worked, hourly wages, annual earnings, occupation, sector and firm. The IVF register contains information on 31,666 women receiving alltogether 96,807 IVF treatments.

C   IVF Treatment

While IVF treatment is the leading medical intervention to help infertile women become pregnant and conceive children, it is often the last in a line of fertility interventions. Women with fertility problems typically visit their general practitioner for medical advice and fertility testing. After a year of having frequent and unprotected sex without getting pregnant, these women are medically diagnosed as infertile and can then be referred to a fertility clinic or hospital if they are below age 40. General practitioners are responsible for

referrals. The Danish National Health Care System entitles women with a referral to have three IVF treatments at no cost. Without a referral, women have to pay.[3]

Once referred to a fertility clinic or hospital, women undergo IVF treatment in four consecutive stages. The first stage involves the intake of fertility medication to stimulate ovaries in the development of eggs. In a normal menstrual cycle, ovaries typically make and release one egg. In a menstrual cycle under medicated stimulation, ovaries make and release several eggs. The second stage involves the collection of these eggs. The third stage involves the actual in vitro fertilization, where eggs and sperm meet under laboratory conditions appropriate for fertilization and early embryo growth. The fourth and final stage involves the implantation of the most promising embryos. IVF treatment can and does often fail at each stage of the treatment: fertility medication may not work, there may be no eggs to retrieve, there may be no suitable embryos to transfer, or the transferred embryos may simply stop growing.

Figure 1 shows how the number of IVF treatments, the IVF success rate, the number of eggs retrieved (second stage), and the number of fertilized eggs transfered back into the womb (fourth stage) evolved in Denmark between the years 1995 and 2005. It illustrates three general findings. First, IVF usage increased until the year 2000, after which usage more or less stabilized. Second, the IVF success rate per treatment increased substantially; in this time period, the IVF success rate nearly doubled. Third, the number of eggs retrieved and the number of embryos implanted gradually fell in the period 1995-2005. The trending up of success rates and trending down of retrieved eggs and transferred embryos we attribute to improvements in IVF technology.

Because the success rate per treatment is still relatively low, most women undergo multiple treatments to achieve success. But not all. After each unsuccessful treatment, there are women who decide to forego further treatment. The share of women who decide to forego further treatment is rising with the number of treatments.[4] With IVF treatments being costly in social, psychological as well as financial terms, these rising shares suggest that the process of

---

[3]Women are also allowed to approach a private clinic instead of waiting for a referral from their general practitioner. This option requires full payment but offers flexibility in terms of the timing of treatment, the number of treatments, and treatments with weaker, if any, age restrictions.

[4]According to the IVF register, about 6 percent of all women stop treatment after a failed first IVF treatment; about 12 percent stop after a failed second IVF treatment; about 24 percent stop after a failed third IVF treatment; and about 30 percent stop after a failed fourth IVF treatment.

IVF treatments is selective and that women who decide to continue treatment are probably women with more resources or a stronger demand for children.

D   IVF Instrument

In order to exploit the IVF process to arrive at the causal link running from fertility to earnings, we need an instrument that somehow captures an exogenous shock in IVF-driven fertility. In constructing the IVF instrument we let the following considerations guide us. Firstly, we know that most women undergo multiple IVF treatments to achieve success. With the number of treatments being endogenous, it does not make much sense to treat the success rate in a sequence of IVF treatments as exogenous. So we focus on first IVF treatments. Secondly, we are interested in how women respond to exogenous variation in fertility measured at the extensive margin. So we consider childless women entering their first IVF treatment. Finally, we want these women to be as similar as can be. So we concentrate on childless women who have successfully reached the fourth stage and had embryos implanted. Within this sample, we choose success at their first IVF treatment as our instrument. For some women, the embryos develop and IVF treatment leads to pregnancy and children; for other women, the embryos stop growing and IVF treatment fails. If the development of implanted embryos in the womb is to a large extent exogenously determined, exogenous treatment success guarantees that after the first full IVF treatment all women are still very similar, except that for some women the IVF treatment has lead to children. In our empirical setup, we will check this claim and test whether the chance of success is somehow related to the pre-treatment labor market characteristics we observe in our data.

E   IVF Sample

From the IVF register, we draw our primary sample: childless women in their first IVF treatment with embryo implants. To do so, we construct three variables for sample selection: treatment order, which is derived from the date of treatment; childlessness, which is derived from the number of children observed the year before the first IVF treatment; and positive embryo implants, which is taken from the number of fertilized eggs transferred back into the womb.

At the outset, we were concerned about measurement error in treatment order. Measurement error may arise because the IVF register does not contain

information on IVF treatments prior to 1994, which is the year the IVF register started. If some women underwent IVF treatment in both 1993 and 1994, for example, we would wrongfully classify the IVF treatment in 1994 as the first one. Measurement error may further complicate estimation because women treated multiple times are possibly different from women treated only once in ways related to preference for children or financial resources. Since few women skip a full year between treatments, we restrict our sample to who began their first treatment in 1995 or later. This should eliminate, or at least reduce, any measurement error and the bias it may entail.

Of the original 31,666 women who were treated at least once sometime between 1994 and 2005, we remove 2,908 women who were treated in 1994, 5,674 women who enter the first IVF treatment with children, 4,286 women who had no eggs inserted, either because of failed egg production or failed fertilization of eggs, and 260 women who could not be matched to the other registers. This leaves us with an IVF sample of 18,538 treated women.

Table 1 provides sample means and standard deviations for some of the pre-treatment characteristics and the post-treatment outcomes. In constructing pre-treatment characteristics, we focus on labor market characteristics observed in the year before treatment. In constructing post-treatment outcomes, we focus on labor market outcomes observed the year of (potential) child birth and later. We do this because we are interested in changes in fertility. We calculate the year of potential child birth by adding 9 months to the day embryos are transfered assuming that women with a failed treatment would have given birth nine months after the first treatment had the treatment been successful. In our IVF sample, the median duration of gestation is 9 months for first successful treatments. We present post-treatment labor market averages taken four years after (potential) childbirth to allow for possible longer-run effects of having children. In our analysis below we follow women up to 10 years after (potential) childbirth. All earnings and wage variables are measured in 2008 Danish Kroner (DDK 100 corresponds to USD 20 in August 2008). Table 1 also provides sample means and standard deviations for a 30 percent representative sample of women born around the same time as IVF treated women. The sample consists of 103,826 women who had their first child somewhere between 1995 and 2005, and thus share their demand for children with the women in our IVF sample. For comparison purposes, the same table further provides differences in sample means between successfully and unsuccessfully treated women, and between first time successfully treated IVF mothers and

9

representative mothers.

We can make two informative comparisons. First, we compare successfully treated women to other women with children. We find that the women in our IVF sample are older, better educated, and more highly paid than the women in the representative sample. These differences, which are all statistically significant, suggest that IVF treated women are on average women with a stronger earnings potential than other women with children. Marianne Bitler and Lucie Schmidt (2012) find comparable differences in the US; that is, older and better educated women are more likely to undergo IVF treatment. Second, we compare women who experience a successful first treatment to women who experience a failed first treatment, before and after the first IVF treatment. Although the successfully treated women in the IVF sample are almost a year younger than the unsuccessfully treated women, they are remarkably similar on almost all pre-treatment labor market characteristics.[5] They attain the same level of education and have exactly the same annual earnings before they seek IVF treatment. They also have partners with similar levels of education and annual earnings. We find that participation rates, as measured by positive annual labor earnings, are statistically higher for successfully treated women than for unsuccessfully treated women, but not in any meaningful way.[6] After the first treatment, however, successfully and unsuccessfully treated women turn out to be systematically different; that is, successfully treated women work fewer hours, earn lower salaries, and are less likely to work for pay. While these fertility and labor patterns hint at a causal relationship between childbearing, labor supply and earnings, we need a more sophisticated analysis to appropriately identify the impact of fertility on labor market outcomes.

---

[5]The observed difference in age at first treatment is not unexpected and accounts for the widely held medical notion that age itself is considered the single most important factor in assisted reproduction (Zev Rosenwaks, Owen Davis and Mark Damario 1995; Allan Templeton et al. 1996; Laura van Loendersloot et al. 2014). In our regression analyses we will estimate fertility effects that take age at first treatment effects into account.

[6]Online Appendix Table 1 reports estimates from regressions linking treatment success (at the first IVF treatment with embryo implant) to pre-treatment labor market outcomes, controlling for year-of-treatment and age-at-first-treatment fixed effects. The estimates attached to the pre-treatment labor market characteristics are all small and, apart from participation, far from statistically significant. Together, these labor market characteristics do not predict treatment success.

III IV Methodology using IVF Treatments

To identify fertility effects at the extensive margin, we make use of a standard IV setup in which success at the first IVF treatment among childless women with embryo implants serves as our instrumental variable; that is, we use a two-stage least squares model to estimate a linear relationship between the annual labor earnings and fertility of woman i who was first treated t years earlier. The first and second-stage regressions are of the following form:

(1) $F_{it} = \alpha_t X_i + \beta_t Z_i + u_{it},$

(2) $Y_{it} = \gamma_t X_i + \delta_t F_{it} + \upsilon_{it}.$

In these two regressions, Y is a measure of female labor earnings;[7] X is a set of exogenous control variables including women's age at first treatment, year of treatment, years of education and pre-treatment labor earnings; F is the endogenous fertility indicator, which equals 1 if a woman has children and 0 otherwise; Z is the instrumental variable, which equals 1 if the first IVF treatment with embryo implants (in a sequence of IVF treatments) has lead to a childbirth and 0 otherwise; and u and υ are the econometric errors, which contain unobservable factors that can either be related to fertility, earnings or both.

The parameter of interest is the second-stage parameter δ_t, which captures the causal effect of fertility (measured at the extensive margin) on labor earnings t years after these women entered their first IVF treatment. This parameter is identified if the IVF instrument satisfies the following three conditions: (a) treatment success is correlated with fertility (relevance); (b) treatment success is more or less randomly assigned (independence); and (c) treatment success exclusively affects labor earnings through its first-stage impact on fertility (exclusion). While the fertility effect (second stage) is the target of estimation, we first examine the effect of IVF treatment success on fertility (first stage) as well as other outcomes (reduced form). We do this for two reasons. First, the first-stage and reduced-form estimates provide insights into the validity of the IVF instrument. Second, the reduced-form estimates are, as unscaled versions of the second-stage estimates, informative about the causal consequences of having children. If we find no reduced-form effect, we know that there is no causal relationship between having children and labor earnings either.

---

[7]In our analysis, we carry out the same fertility regressions for other female labor market outcomes, including whether women work for pay, their hours worked, hourly earnings and what jobs they hold. We also consider their partner's earnings.

IV Reduced-Form Results: IVF Treatment Effects

In this section we estimate how IVF treatment success (at the first IVF treatment with embryo implants) affects several outcomes, including childbearing (first stage) and annual earnings, labor supply and wages (reduced form). We present the IVF treatment effect estimates in event study graphs and tables. The graphs visualize the relationship between IVF treatment success, childbearing and labor market outcomes after taking out years of education, year of treatment and age at first treatment effects, for six years preceding the year of (potential) childbirth, the year these women have their (potential) first child (which we refer to as year zero), and for the 10 years following (potential) childbirth. The treatment effect estimates in the event studies are normalized to 0 in period -1. The tables report the corresponding IVF treatment effect estimates for short, medium, and long-run outcomes, which we construct as outcome averages over the years 0 to 1, 2 to 5 and 6 to 10. We also provide treatment effect estimates for the pre-treatment baseline mean, which we construct as the outcome average taken over a four year period prior to treatment entry. If IVF treatment success is independent of pre-treatment outcomes, these estimates should be zero. To gain further precision, we control for years of education, year of treatment, age at first treatment, and the pre-treatment baseline outcome mean (in case of post-treatment outcome regressions). Pre-treatment baseline means and sample sizes are also shown.

A Childbearing

Does IVF treatment success at first IVF treatment with embryo implants predict fertility, measured at the extensive margin? Given that IVF treatment is the leading medical intervention to help infertile women get pregnant and conceive children, the answer is likely affirmative. Figure 2a provides graphical evidence demonstrating that success is indeed strongly correlated with fertility. We see that treatment success has a strong and immediate impact on the likelihood of having children. In year zero we do not see an impact of 1 because most women after a first failed treatment continue treatment the same year. Some of these later treatments are successful. In subsequent years, we see that the impact of the treatment is falling, but always remains positive, suggesting that our IVF instrument has predictive power; that is, childless women whose

first IVF treatment with embryo implants did not lead to pregnancy and childbirth are also more likely to remain childless in the medium and longer run. The first-stage estimates, which we report in column 1 of Table 2, confirm that success at first IVF treatment predicts a long-lasting difference in the chance of having children. All the treatment effects we estimate are positive and statistically significant (with corresponding F statistics far beyond the typical rules of thumb values for instrument relevance).

B   Labor Market Outcomes

Does IVF treatment success affect the labor market career of women? Figure 2b shows the relationship between treatment success and female annual labor earnings, which is our main labor market outcome. We see that the impact of a successful IVF treatment is virtually zero before (potential) childbirth, suggesting that treatment success does not depend on the earnings women receive before they start their IVF treatment. We also see that a successful IVF treatment has a large negative impact on female annual earnings at the time of (potential) childbirth. The negative impact is the largest during the first two years, starts to decline thereafter but is still there when first-born children are about ten years old. The corresponding estimates in column 2 of Table 2 are all negative and statistically significant in the short, medium, and long run. The estimated effect on pre-treatment annual earnings as outcome is practically zero and statistically insignificant, suggesting that treatment success at first IVF treatment with embryo implants is as good as random (conditional upon education, IVF treatment year, age of the mother at the first treatment fixed effects). Our estimates thus imply that these women earn persistently less because of childbearing and not because of something else.

Female annual labor earnings depends on the number of hours women work in a year as well as the hourly wage rate. To further examine the source of the observed fall in annual labor earnings, we also look at the treatment impact on female labor force participation, the number of hours worked, and wage rates. Figures 2c, 2d and 2e are the event studies for these alternative labor market outcomes. We first take a look at labor supply. If the fall in annual earnings is driven by women working less, we should see that successfully treated women are more likely to stop working or work fewer hours (when they work). Figures 2c and 2d show that women are more inclined to stay at home or work fewer hours shortly after a successful first treatment, but return to the labor market

three or four years after treatment and then continue to participate as much and work as many hours as women with a failed first treatment. We next consider the wage rate. If the persistent fall in annual earnings is not driven by women working less, we should see that women get lower hourly wages, at least in the longer run. Figure 2e confirms this. We find little effect on the wage rate during the first two years following a successful treatment, but negative effects in the long run. In columns 5 and 6, we find that hourly wages fall with 2 to 3 percent in the medium and long run.[8]

We can also examine how the annual earnings of partners of IVF treated women respond to the IVF treatment. We focus on partners in couples at the time of treatment. Figure 2f plots the relationship between treatment success and male annual labor earnings. In contrast to the earnings results for women, we do not see much for their partners. If there is a labor response, it seems that having a successful IVF treatment reduces their annual earnings. In column 7 the estimated short and medium-run effects of IVF treatment success on partner's annual earnings are small, and much smaller than those estimated for women. Only the short-run effect is estimated precisely enough to be statistically significant. The long-run effect appears somewhat smaller than the long-run effect for women, but is too imprecisely estimated to be informative.

An equally important question is whether IVF treatment success is (somehow) related to the labor market outcomes of women observed in their pre-treatment years. We find no evidence of this. When we plot the event study graphs, all the trends in the pre-treatment period are flat. When we regress treatment success on a set of pre-treatment labor market outcome averages, all our treatment effect estimates are close to zero and far from statistically

---

[8]Danish registers hold records on annual earnings, full-time employment status, monthly hours worked and hourly wages. Hourly wages are computed by dividing annual earnings by annual hours worked, which is an aggregate of monthly hours worked. Because monthly hours are recorded for one month a year for workers who earn more than 10,000 DKK (measured in 2008 DKK), we do not observe monthly hours worked and hourly wages for all workers. The restricted sample contains 14,022 observations. Of the original 18,538 observations, we observe 16,996 women with positive annual earnings. Among those, we lose 1,428 women with no records on hours and wages, 1,300 women with no records on pre-treatment hours and wages, and another 246 women who either work more than 13 hours per day or earn less than 50 DKK per hour. We recognize that the regression results for the number of hours worked, hourly earnings (and log hourly earnings) are obtained on smaller and possibly selective samples. In our sensitivity section we test whether sample selectivity is affecting our regression results. We find no evidence that the estimated fertility effects on labor supply and hourly earnings are tainted by sample selection bias.

significant.[9]

C Depression and Divorce

Can IVF treatment success influence labor earnings for reasons other than childbearing? We consider depression and divorce as two possible pathways. In particular, we think that disappointment after a failed IVF treatment can ultimately lead to depression and divorce, possibly biasing results. Women who are depressed may work less if depression, or severe symptoms thereof, interferes with their ability to work. Women who divorce (or anticipate divorce) may work and earn more to compensate for the loss in spousal income (Kelly Bedard and Olivier Deschênes 2005; Betsey Stevenson and Justin Wolfers 2007; and Olivier Bargain et al. 2012). We augment previous labor market results and use depression (measured as antidepressant medication use) and divorce as additional outcome variables.

We first check whether IVF treated women are more likely to get depressed (and subsequently work less) after a failed treatment. Figure 2g provides little evidence of this. If there is an impact, it is only small. The corresponding estimates in column 8 confirm this. We next check whether divorce rates increase after a failed treatment among the IVF treated women in our sample. Figure 2h does not show such a divorce pattern, at least not in the long run. In the short and intermediate run, we estimate a modest but statistically significant decline in divorce rates of 1 percentage point among successfully treated women. In the long run, however, the estimated impact of a successful treatment on divorce is slightly positive, quite small, and statistically insignificant. We conclude that women who seek IVF treatment face similar long-run divorce risks, independent of treatment success.[10]

D Some Sensitivity Checks

Our treatment effect estimates may be subject to a number of biases: omitted variable bias related to female health factors (among others), sample selection bias related to wages being observed only for women who work, and sample

---

[9]One exception is the estimated treatment effect on labor market participation history, which is small but statistically significant. This occurs -we think- more or less by chance. Intuitively, we expect to find a statistically significant association between treatment success and labor market history in one out of 20 specifications.

[10]In an accompanying paper, we examine how children influence marital stability in much more detail (Petter Lundborg, Erik Plug and Astrid Würtz Rasmussen, 2016).

selection bias related to the unbalanced nature of our short, medium and long-run samples. To make sure that our findings are not driven by these biases, we perform several sensitivity checks. Online Appendix Table 2 contains these sensitivity results.

One concern is that unsuccessfully treated IVF women may be less healthy and, because of that, work fewer hours (in the short run) and receive lower salaries (in the long run). To address this, we run two regressions that are arguably informative about unhealthy women, treatment success and earnings. One regression includes additional controls for the number of eggs collected and transferred, diagnoses, causes of infertility, type of IVF treatment, and clinic indicators, which could all reflect potential health risks at the time of treatment. The estimates in column 1 remain practically unaffected. Another regression is run on a restricted sample of a priori healthier women, which we define as those IVF treated women whose infertility problem is on their partner's side. Again, the estimates in column 2 do not change much when we move to a sample of healthier women; if anything, the estimates are somewhat larger in magnitude but less precise, due to the smaller sample size.[11]

Another concern is that the treatment effect estimates for hours worked and hourly wages are obtained for IVF treated women for whom we observe hours and hourly wages. To address the issue of endogenous sample selection, we examine whether the missing information on hours and hourly wages creates any sample composition bias by estimating the effect of child birth on pre-treatment education and earnings in our sample of workers. In columns 3 and 4, we show that these estimates are insignificant, suggesting that child birth does not seem to create any important sample composition bias in terms of pre-treatment earnings and education. Since sample selection occurs independent of the treatment, at least in the medium and long run, the fertility effects we find for medium and long-tun labor supply and hourly earnings are unlikely tainted by sample selection bias.[12]

A final concern is that the unbalanced panel structure, which arises because information on long run labor market earnings is not available for women who

---

[11]With comparable treatment effects for healthier women, we also dismiss a related concern that IVF treatments carry heterogeneous (mental) health risks; that is, treatment induces larger (mental) health risks among unsuccessfully treated women.

[12]Our results are further insensitive to including women with missing data on pre-treatment wages and hours and not controlling for these factors in the regressions. The effects of having children on hourly wages are then 7.20, -25.08, and -21.71 in the short, medium, and long run. If we further remove the restrictions on hourly wages and hours, the point estimates only slightly change to 7.33, -24.37, and -22.73.

recently entered IVF treatment, may cause sample selection bias in our labor response estimates. We find no evidence of this. The earnings estimates obtained on a balanced sample (reported in column 5) are identical to those reported for the unbalanced sample.

V Main Results: Fertility Effects at the Extensive Margin

A Labor Market Outcomes

Table 3 contains the IV estimates of the effect of having children on the labor market outcomes of women, based on the second-stage specification in equation (2). In column 1, we take annual labor earnings as our labor market outcome, fertility measured at the extensive margin as our endogenous variable, success at first IVF treatment as our instrument and IVF treatment year, the age of the mother at the first treatment, education and pre-treatment earnings as control variables. In subsequent columns, we turn to labor market participation, hours worked, and hourly wages as labor market outcomes. As before, we replace the pre-treatment labor market control variable with the pre-treatment values of the labor market measure under study. We calculate averages of the labor market outcome variables over the pre-treatment years -4 to -1 and post-treatment years 0 to 1, 2 to 5, and 6 to 10. We also present the fertility effect estimates in terms of percent impacts which we define as the ratio between the IV estimates and sample means of the corresponding labor market outcomes in the short, medium and long run.

In column 1 we find that having children reduces earnings by DKK 70,000 in the short run, DKK 30,000 in the medium run, and DKK 30,000 in the long run. We attribute the larger earnings effect observed in the short run to mothers taking up maternity leave for an extended period at a lower compensation rate. In the medium and long run, when most mothers have finished their maternity leave, we find that the estimated effects decline but remain negative and statistically significant. In terms of percentage impact, the medium and long-run estimates imply that Danish women earn about 11-12 percent less because of children. The consequences of children, measured at the extensive margin, are therefore substantial.

In columns 2 and 3 we also find that women work less because of children,

but only when children are young. In the short run, participation declines 7 percentage points. And for those women with positive earnings, hours fall on average 6 hours per week (which is a labor supply reduction of 21 percent). In the long run, however, participation rates and hours worked are unaffected by having children. All the long-run labor supply estimates are statistically insignificant and practically zero. Interestingly, the estimates for hourly wages in column 4 show the opposite pattern. We find little effect on the wage rate during the first two years following the (potential) birth of the child, but significant, negative, and large effects in the medium and long run. In particular, hourly wages fall by 11-13 percent in the medium and long run.

We therefore conclude that the labor market consequences of having children are large for women. When children are young, women earn less because they work less. When children are older, women earn less because they get lower wages.[13]

B Possible pathways

We next explore possible pathways for the decline in earnings. In particular, we examine whether differences in earnings can be explained by the differences in job characteristics of the jobs held by women with and without children. To this end, we use matched employer-employee level data and test whether women make different job choices because of children in terms of job changes and job characteristics. To measure job changes, we use dummy variables indicating whether any workplace and (four-digit) occupation change took place in the short, medium, and long run. To measure job characteristics, we take averages of earnings, hours, and hourly wages (in logarithms) by workplace and (four-digit) occupation over the entire study period from the representative sample of Danish women. If women with children sort into part-time and lower-paid jobs, the latter outcomes should indicate that having children leads to a reduction in the workplace-based averages of hours worked and hourly earnings. We use the same approach to examine whether women sort into

---

[13]The interpretation of IV estimates gets more complicated when most women after a failed first treatment end up having children. In Online Appendix C, we show that delayed fertility may bias our fertility effect estimates downwards. In particular, we show that the medium-run effects of having any children on work effort and earnings are likely underestimated. Apart from the small but positive impact we estimate for hours worked, the medium-run effects of having children on earnings, participation and wages are all highly significant and negative. Hence, we conclude that the possible impact of delayed fertility is not strong enough to overturn our main conclusion.

public sector jobs, typical female jobs (measured as the percent female workers for each workplace averaged over our study period), or jobs closer to home. Table 4 contains the estimates on job changes and job characteristics.

In columns 1 and 2 we find that women with children more frequently change occupation and workplace than women without children. These positive mobility effects are mostly observed in the short and medium run, with estimates only being significant in the short run. In columns 3 and 6 we find that women who change occupation and workplace because of children also sort into occupations and workplaces with on average lower earnings. In addition, we find that once these women start working in occupations and workplaces that pay less, they keep on working in such occupations and workplaces. The long-run estimate for the occupation earnings average suggests that women, 6 to 10 years after giving birth, work in occupations in which the average female worker earns approximately DKK 5,000 less, which is about 2 percent of the occupation earnings average in the sample.[14] Although the estimates are all small and not always precise enough to be statistically significant, the negative effect of having children on average occupation and workplace earnings seems to persist or even to rise in the medium and long run.

In the remaining columns we explore, in more detail, the type of lower-paid jobs women hold after having children. In most cases, our estimates are insignificant and too small; that is, we find no evidence that women, after having children, sort into part-time jobs, public sector jobs, or more typical female jobs. One job characteristic, however, stands out: distance between home and work. In the short run, we find that women with children work somewhat further away from home than women without children, which we attribute to women moving to the suburbs after having children. In the medium and long run, however, our estimates indicate that women with children work much closer to home than women without children. The medium and long-run effects of having children on distance between home and work are substantial. For example, our long-run estimate of -5.53 kilometer implies a 45 percentage point reduction in the average work-related commuting distance. In light of these

---

[14]The estimates are small but not implausible. When interpreting these estimates one should keep in mind that by using occupation and workplace earnings averages as outcomes, there will be much less earnings variation in the data. This is even more so in countries with more compressed earnings distributions, such as Denmark. It is therefore helpful to convert these effects into units of standard deviations in the full distribution of occupation and workplace earnings. When we do so we find more plausible effects; in the medium and long-run, the effects of having children on average occupation earnings amount to about 0.1 of a standard deviation.

findings, we believe that women with children reveal a stronger preference for working close to home than women without children, and as a consequence end up working in less well paid jobs.

C Heterogeneity

So far we have focused on the average impact of having children. To examine whether the labor market consequences of having children are different for different women, we estimate fertility effects on different sample splits using pre-treatment indicator variables for high earnings, for college education, for ages older than 32 years old, for high partner earnings, for public sector jobs, and for being treated in 2002 or later. We refer to the corresponding sample splits as reference group and indicator group samples. Table 5 presents these estimates for these samples; the top panel contains fertility effect estimates for the reference group samples, the middle panel contains fertility effect estimates for the indicator group samples, and the bottom panel tests whether the fertility estimates reported in panels A and B are different.

We first examine whether fertility effects vary with the earnings potential of women or their partners. These are relevant margins if a higher value of time also raises the cost of having children; the value of women's time is, in theory at least, linked to their own wage rate and that of their partner (Reuben Gronau 1973; Becker 1981). We proxy the earnings potential of women with pre-treatment earnings, education and age-at-treatment. Our estimates suggest that women with high earnings potential are more labor market responsive to having children than women with low earnings potential. For the women in our reference samples, being less than 32 years old, having pre-treatment earnings in the bottom 75 percent of the earnings distribution and having no college degree, the estimated fertility effects in panel A indicate that having children reduces earnings by 33 to 35 percent in the short run, 8 to 12 percent in medium run, and 6 to 12 percent in the long run. For the high-potential women in the indicator group samples, we find that women in the top 25 percent of the pre-treatment earnings distribution reduce their earnings substantially more than do other women. In the medium and long run, these fertility effect estimates are much larger in both absolute and relative terms. In column 2 we find (somewhat surprisingly) that having a college degree, which we define as 15 or more years of schooling, does not affect the impact

on annual earnings.[15] In column 3 we find differences by age at treatment; that is, women older than 32 reduce their annual labor earnings more than younger women in the short, medium and long run. In column 4 we examine whether fertility effects vary with the earnings potential of partners. Here we see the largest differences. Women with partners in the top 25 percent of the male pre-treatment earnings distribution reduce their earnings significantly more that women with low-potential partners in the short, medium and long run. After having children, the fertility effects are large and amount to, what seems, a structural reduction of approximately 30 percent in annual earnings.

As a further check, we examine how earnings responses vary across the pre-treatment earnings distribution of women and their partners. In constructing graphs, we create percentile ranks of pre-treatment labor earnings and plot the fertility effect estimates taken from IV regressions run on samples restricted to mutually exclusive percentile ranks 1-20, 21-40, 41-60, 61-80 and 81-100. Figure 3a shows earnings responses by pre-treatment earnings of women. Observe that all earnings responses represent earnings losses, regardless of earnings potential and period. In the short run, we see that earnings losses modestly increase in pre-treatment earnings. In the medium and long run, we see that earnings losses are larger for women in the second and fifth quintile of the earnings distribution. Figure 3b shows earnings responses by pre-treatment earnings of partners. The earnings response patterns are rather similar. In the short run, we see again that women experience earnings losses that modestly increase in the pre-treatment earnings of their partners. And in the medium and long run, we see that earnings losses are more or less the same at the bottom half of the distribution and get much larger in the upper half of the distribution. It seems that women at the top of their own or partner's earnings distribution experience the largest earnings losses because of having children.

We also investigate whether the fertility effects vary with public sector jobs and time. The reason for showing these interactions is that maternity leave arrangements are more generous in public sector jobs as well as after 2002. Women in public sector jobs on paid maternity leave receive a 100 percent compensation rate (which is part of our annual labor earnings measure).

---

[15]We have experimented with years of education interactions and probed sample splits by having compulsory education, having no college, having some college education and having a college degree. The latter specification is the one we report. The main message is that we do not find much when we allow the fertility effects to vary by education. Such effects could arise if women in Denmark face relatively low returns to education. Colm Harmon, Ian Walker and Niels Westergaard-Nielsen (2001) provide a cross country analysis of the returns to education in Europe and find one of the lowest returns for women in Denmark.

Women in private sector jobs on paid maternity leave receive less. In column 5 the fertility effect estimates confirm that women in public sector jobs experience a smaller earnings loss than women in private sector jobs. This difference is statistically significant, but only in the short run when most women take up their paid maternity leave. In 2002 Denmark expanded their parental leave arrangements. Women could take up 4 extra weeks of paid maternity leave and 22 extra weeks of shared parental leave. In column 6 we find that women with children respond to the reform by reducing their labor earnings, but only in the short run.

VI Other Results: Fertility Effects at the Intensive Margin

One objective of this paper is to distinguish extensive from intensive fertility margins. In particular, we want to know whether the effect of children on the labor market outcomes of women is larger at the extensive margin than at the intensive margin, just as simple theories predict.

A IVF Treatments and Twins

IVF treatments also prove helpful in generating natural experiments with independent variation in fertility measured at the intensive margin. First, we can apply our IV strategy using success at the first IVF treatment as an instrument on a sample of IVF treated women who already have children when they start their treatment. The instrument generates variation in fertility, similar to the variation we use for childless women, except that it raises the likelihood of going from one to two or more children for mothers who enter treatment with one child. Second, we can exploit the larger fraction of twin births among IVF births (about 23 percent of all successful treatments result in twins) and apply the more commonly used IV strategy using twins at first birth as an instrument on a sample of successfully treated women who received at least two embryo implants. In this case, we zoom in on IVF treated women (with two or more embryo implants) whose IVF treatment resulted in at least one child, assume that a successful development of one, two or more implanted embryos is to a large degree exogenously determined, and use having twins as an instrument to raise the likelihood of going from one to two (or more)

children. Online Appendix D Table 4 shows descriptive statistics for these subsamples.

Table 6 contains IV estimates of fertility effects on annual earnings, measured at both extensive and intensive margins. In column 1 we reproduce the IV fertility estimates at the extensive margin for the sample of childless women entering their first IVF treatment (our baseline sample). In columns 2 and 3 we report IV fertility estimates at the intensive margin. In column 2 we show fertility estimates that come from IV estimation using the first IVF treatment as an instrument on a sample of IVF treated women with children. In the short run, we observe that such mothers experience substantial and significant earnings losses, albeit somewhat smaller in size than the earnings losses observed at the extensive margin. In the medium run, the earnings losses get much smaller, but remain precise enough to be statistically significant. And in the long run, the earnings losses are gone. If anything, we estimate small and insignificant earnings gains.

In column 3 we turn to fertility estimates using twins at first birth as an instrument on a sample IVF treated women with twins and singletons. With the twin instrument, we find that women earn less (because of children) in the short, medium and long run. These earnings losses are substantially smaller than the corresponding earnings losses observed at the extensive margin (column 1). The earnings losses are also substantially smaller than the corresponding earnings losses we estimate at the intensive margin with the treatment success instrument (column 2), but only in the short run. This is because most mothers take up full maternity leave after childbearing, regardless of giving birth to twins. The differences between the medium and long-run estimates are not large enough to be statistically significant.[16]

All these estimates indicate that the labor market consequences of childbearing, measured at the intensive margin, are relatively small and mostly short lived. Taken together, these results provide clear evidence that fertility effects are much stronger at the extensive margin than at the intensive margin.

---

[16]At the bottom of Table 6, we also report the estimated effect of the instrument on pre-treatment earnings. The corresponding reduced-form estimate should be zero. With the treatment success instrument, the estimate is small but positive and statistically significant at the 10 percent level (the amount is only 3 percent of the average earnings of women in this treatment success instrument sample). With the twin instrument, the reduced-form estimate is equally small but negative and statistically insignificant (the amount is only 2 percent of the average earnings of women in this twin instrument samples).

B External Validity

It is quite clear that women who decide to enter IVF treatment are different from a larger population of representative women; they are better educated, work more, earn higher salaries, show an explicit demand for children, and are older when they have children. If these observable differences also mean that their labor market responses to having children are different, it is natural to ask what we can learn from a sample of IVF treated women.

One way to examine the wider generalizability of IVF fertility findings is to expose IVF treated women and other women, not treated with IVF, to the same natural experiment and compare their labor earnings. The natural experiment we have in mind is the twin experiment, where twins at first birth generate fertility variation at the intensive margin. We conjecture that with comparable labor supply responses, IVF results are generalizable. Table 6 also reports fertility estimates based on the twin experiment for the representative sample of Danish women. In column 4 we find that the fertility estimates are remarkably similar to those obtained for IVF women, shown in column 3, and particularly so in the short run. Among IVF mothers, one additional child reduces earnings by DKK 14,500, whereas the corresponding effect among other mothers is DKK 13,000. In the medium run, earnings are reduced by DKK 6,000 and DKK 3,800, respectively. We also find that fertility estimates are statistically similar in the longer run, but we readily confess that the estimates are rather imprecisely estimated.[17]

C Comparison to US Twin Births

While our IVF-based fertility results may generalize to other Danish women (not treated with IVF), they may be less informative about women having children in other developed countries with less family-friendly policies. Here we try to get some mileage from comparing the twin-based fertility results reported in Table 6 with other twin-based fertility results taken from the US, which is one of the few developed countries that does not guarantee paid

---

[17]In Online Appendix Table 5 we also compare labor earnings of women in our baseline IVF sample to those of women in a more representative sample drawn from the full population of women who had their first born child around the same time as IVF treated women had their first IVF attempt. In this Appendix, we also show that conditional and unconditional measures of labor earnings in the IVF and representative samples follow roughly the same pattern before and after childbirth, which is suggestive that any inherent differences between these two groups of mothers are probably not the leading cause of any differences in how they respond to first born children.

maternity leave to female workers.

We have surveyed three US studies on the impact of children on female labor market outcomes using twins at first birth as instrumental variable (Bronars and Grogger 1994; Jacobsen et al. 1999; Vere 2011). When we focus on annual labor earnings as the main labor market outcome, select impact estimates from the 1980 (or more recent) Census Public Use Micro Samples (PUMS), and convert each estimate into percentage point changes in average annual earnings to facilitate the comparison, we find that even the smallest of all these estimates (obtained with US samples) is larger than the ones we obtain using a representative sample of all Danish women. The smallest of the fertility effect estimates reported in these US studies is -5 percentage points for having a second child, regardless of the child's age (we get -3 percentage points), and -22 percentage points for having a second child when the child is 2 years old or younger (we get -8 percentage points).[18]

VII   Concluding remarks

This paper evaluates how fertility choices made at the extensive and intensive margins affect the labor market career of women. To do so, we introduce a novel IV strategy based on IVF induced fertility variation using the census of IVF treated women in Denmark. Because observed chances of IVF success do not depend on the labor market histories of women before they enter the IVF treatment, success on the first IVF attempt provides a plausible instrument for childbearing among women without children, as well as women with children.

Our findings help us to answer the questions we posed earlier. First, does having children hurt women's careers (the extensive margin)? They clearly do. The fertility effects we estimate at the extensive margin are negative, large and long lasting. When children are young, we find lower annual earnings because women who would otherwise work decide to work fewer hours, or stop working for pay altogether. When children get older, however, we find lower annual earnings because women receive lower hourly earnings. In particular, we find that women who would otherwise work in better-paid jobs move into lower-paid jobs that are much closer to home. Second, do women earn less because of second born (or later born) children (the intensive margin)? The answer is yes and no. In the short run, the fertility effects we estimate at the

---

[18]In Online Appendix F we discuss in more detail how we arrive at these fertility effect estimates.

intensive margin are negative, albeit more modest in size. In the long run, these negative fertility effects fade out and disappear. Together, our findings leave little doubt that fertility effects are much stronger at the extensive margin than at the intensive margin.

While our findings confirm the widely held view that the labor market consequences of childbearing are substantial and long lasting, it is important to remember that we are studying the impact of children on labor market careers of IVF treated women in Denmark, where parents have access to generous maternity leave, job protection and child care benefits. In particular, questions arise about the wider generalizability of these IVF findings. First, are the labor market effects we find for IVF treated women generalizable to other women? They probably are. We find that IVF treated women respond to exogenous fertility shocks at the intensive margin in similar ways as do other Danish women. Second, how do women respond to having children in developed countries that have less generous maternity leave arrangements? With our data at hand, we can not provide an answer. We can however speculate. On the one hand, we may find smaller fertility effects if women who would otherwise take more time away from work take up less maternity leave when such leave is not readily available. On the other hand, we may find larger fertility effects if women who would otherwise return to their employer at the end of their maternity leave the labor market altogether. We provide some suggestive evidence that the adverse labor market consequences of having children are stronger in developed countries with less family-friendly policies. With Denmark being one of the most generous countries in terms of maternity leave arrangements, our IVF estimates then turn into conservative estimates that are informative about the causal impact of having children on labor market outcomes for women facing less generous maternity leave arrangements.

References

Adda, Jérôme, Christian Dustmann, and Katrien Stevens, "The Career Costs of Children," *Journal of Political Economy*, forthcoming.

Agüero, Jorge and Mindy Marks, "Motherhood and Female Labor Supply in the Developing World Evidence from Infertility Shocks," *Journal of Human Resources*, 2011, *46* (4), 800–826.

Angelov, Nikolay and Arizo Karimi, "Mothers' Income Recovery after Childbearing," IFAU Working Papers 20, Institute for Evaluation of Labour Market and Education Policy October 2012.

Angrist, Joshua and William Evans, "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review*, June 1998, *88* (3), 450–77.

Bargain, Olivier, Libertad González, Claire Keane, and Berkay Özcan, "Female Labor Supply and Divorce: New Evidence from Ireland," *European Economic Review*, 2012, *56* (8), 1675–1691.

Becker, Gary, *A Treatise on the Family*, Harvard university press, 1981.

Bedard, Kelly and Olivier Deschênes, "Sex Preferences, Marital Dissolution, and the Economic Status of Women," *Journal of Human Resources*, 2005, *40* (2), 411–434.

Bertrand, Marianne, "New Perspectives on Gender," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4B, Elsevier, 2011, pp. 1543 – 1590.

Bitler, Marianne and Lucie Schmidt, "Utilization of Infertility Treatments: The Effects of Insurance Mandates," *Demography*, 2012, *49* (1), 125–149.

Bronars, Stephen and Jeff Grogger, "The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment," *American Economic Review*, December 1994, *84* (5), 1141–56.

Browning, Martin, "Children and Household Economic Behavior," *Journal of Economic Literature*, 1992, pp. 1434–1475.

Cristia, Julian, "The Effect of a First Child on Female Labor Supply: Evidence from Women Seeking Fertility Services," *Journal of Human Resources*, 2008, *43* (3), 487–510.

Cruces, Guillermo and Sebastian Galiani, "Fertility and Female Labor Supply in Latin America: New Causal Evidence," *Labour Economics*, June 2007, *14* (3), 565–573.

Feyrer, James, Bruce Sacerdote, and Ariel Stern, "Will the Stork Return to Europe and Japan? Understanding Fertility within Developed Nations," *Journal of Economic Perspectives*, Summer 2008, *22* (3), 3–22.

Goldin, Claudia, *Understanding the Gender Gap: An Economic History of American Women*, Oxford University Press, 1992.

Gronau, Reuben, "The Effect of Children on the Housewife's Value of Time," *Journal of Political Economy*, Part II, 1973, *81* (2), S168–99.

Harmon, Colm, Ian Walker, and Niels Westergaard-Nielsen, *Education and earnings in Europe: a cross country analysis of the returns to education*, Edward Elgar Publishing, 2001.

Hirvonen, Lalaina, "The Effect of Children on Earnings using Exogenous Variation in Family Size: Swedish Evidence," Swedish Institute for Social Research Working Paper, University of Stockholm 2009.

Iacovou, Maria, "Fertility and Female Labour Supply," Institute for Social and Economic Research Working Paper, University of Essex 2001.

Imbens, Guido and Joshua Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, March 1994, *62* (2), 467–75.

Jacobsen, Joyce, James Wishart Pearce, and Joshua Rosenbloom, "The Effects of Childbearing on Married Women's Labor Supply and Earnings: Using Twin Births as a Natural Experiment," *Journal of Human Resources*, 1999, *34* (3), 449–474.

Kleven, Henrik, Camille Landais, and Jacob Søgaard, "Children and Gender Inequality: Evidence from Denmark," Unpublished Manuscript, London School of Economics 2016.

Lundborg, Petter, Erik Plug, and Astrid Würtz Rasmussen, "Children and Divorce," Unpublished Manuscript, University of Amsterdam 2016.

Maurin, Eric and Julie Moschion, "The Social Multiplier and Labor Market Participation of Mothers," *American Economic Journal: Applied Economics*, January 2009, *1* (1), 251–72.

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

Nielsen, Helena Skyt, Marianne Simonsen, and Mette Verner, "Does the Gap in Family-Friendly Policies Drive the Family Gap?," *Scandinavian Journal of Economics*, 2004, *106* (4), 721–744.

Pylkkänen, Elina and Nina Smith, "The Impact of Family-Friendly Policies in Denmark and Sweden on Mothers' Career Interruptions Due to Childbirth," *IZA Discussion Paper No. 1050*, 2004.

Rosenwaks, Zev, Owen Davis, and Mark Damario, "The Role of Maternal Age in Assisted Reproduction.," *Human Reproduction*, 1995, *10*, 165–173.

Rosenzweig, Mark and Kenneth Wolpin, "Life-Cycle Labor Supply and Fertility: Causal Inferences from Household Models," *Journal of Political Economy*, April 1980, *88* (2), 328–48.

Simonsen, Marianne, "Price of High-Quality Daycare and Female Employment," *The Scandinavian Journal of Economics*, 2010, *112* (3), 570–594.

Stevenson, Betsey and Justin Wolfers, "Marriage and Divorce: Changes and their Driving Forces," *Journal of Economic Perspectives*, Spring 2007, *21* (2), 27–52.

Templeton, Allan, Joan Morris, and William Parslow, "Factors that Affect Outcome of In-Vitro Fertilisation Treatment," *The Lancet*, 1996, *348* (9039), 1402–1406.

van Loendersloot, Laura, Sjoerd Repping, Patrick Bossuyt, Fulco van der Veen, and Madelon van Wely, "Prediction Models in *In Vitro* Fertilization; Where Are We? A Mini Review," *Journal of Advanced Research*, 2014, *5* (3), 295–301.

Vere, James P, "Fertility and Parents Labour Supply: New Evidence from US Census Data," *Oxford Economic Papers*, 2011, *63* (2), 211–231.

Waldfogel, Jane, "Understanding the "Family Gap" in Pay for Women with Children," *The Journal of Economic Perspectives*, 1998, *12* (1), pp. 137–156.

Weiss, Yoram, "The Formation and Dissolution of Families: Why Marry? Who Marries Whom? And What Happens upon Divorce," *Handbook of Population and Family Economics*, 1997, *1*, 81–123.

Tables and Figures
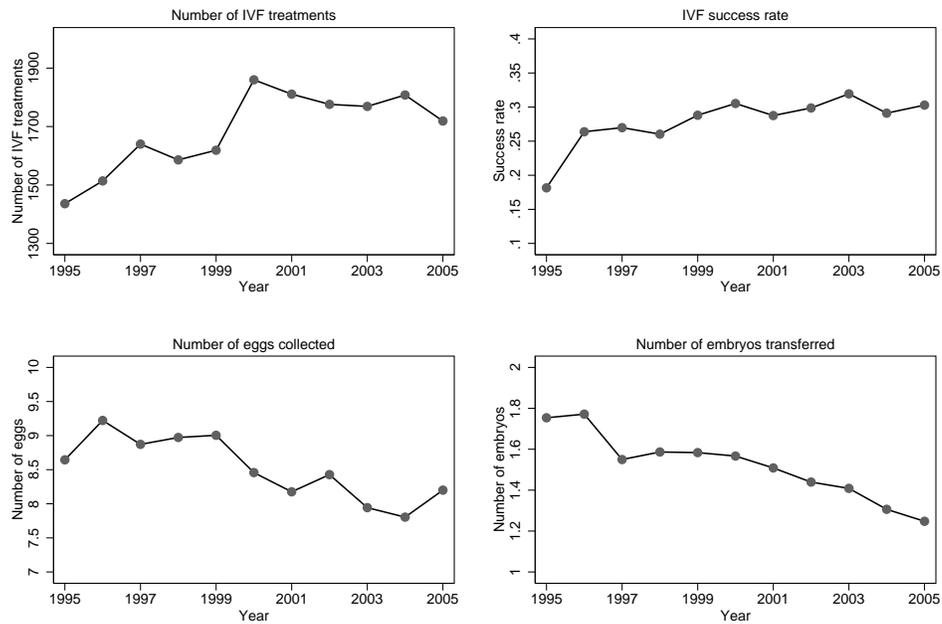
Figure 1: IVF Treatments and Technology 1995-2005

Notes: The figure plots mean values between 1995-2005. The means are calculated for the sample of IVF women who entered their first treatment in 1995-2005 with no children.
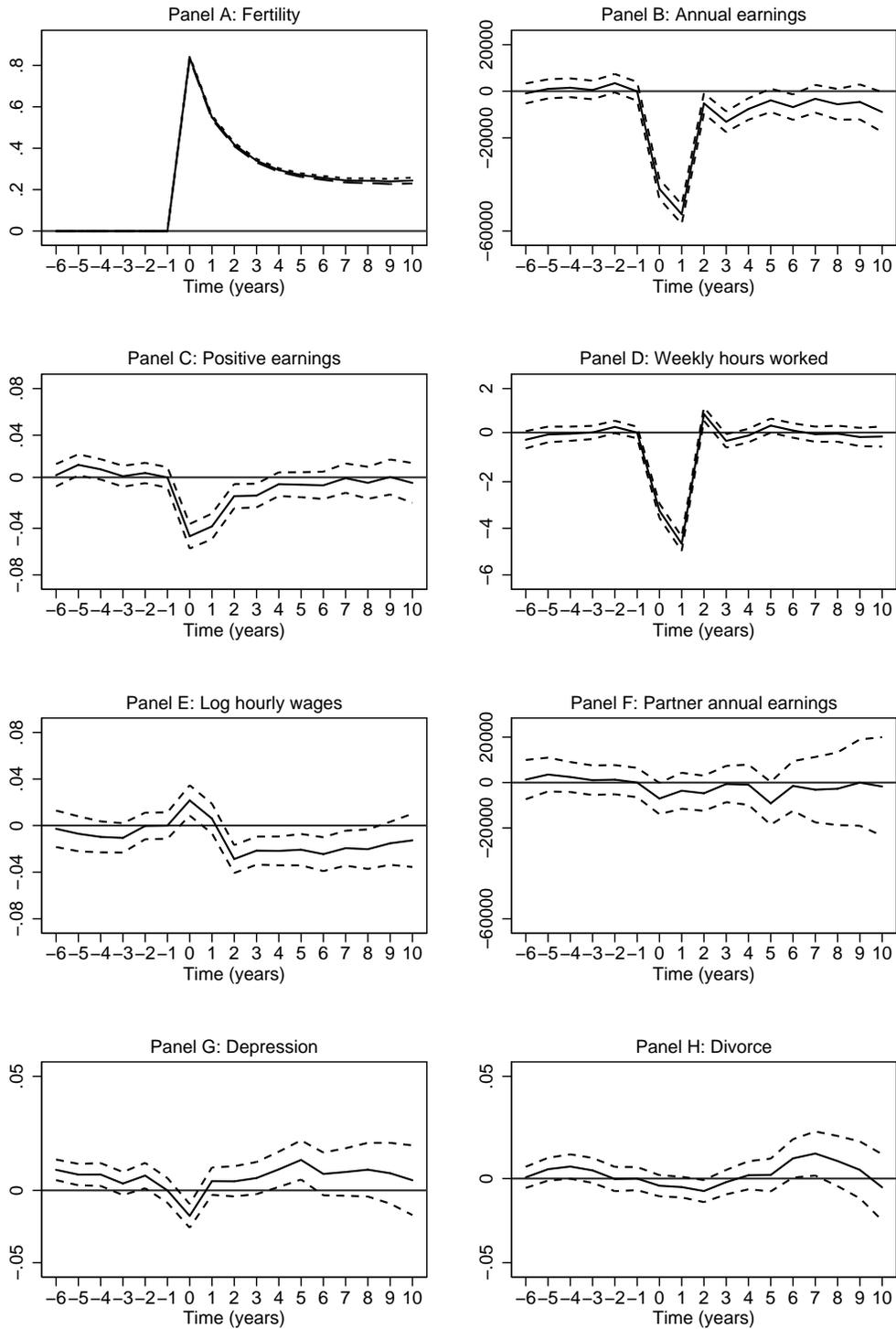
Figure 2: Event Study Graphs of IVF Treatment Effects on Various Outcomes

Notes: The figures plot coefficients from an event-study analysis. Event time is defined as years before and after (potential) child birth. The coefficients in period -1 are normalized to 0. The models are estimated for the sample of IVF treated women who had their first IVF treatment between 1995-2005.The coefficients in panel F are estimated for the partners' of the IVF treated women. For details on the model, see the text.
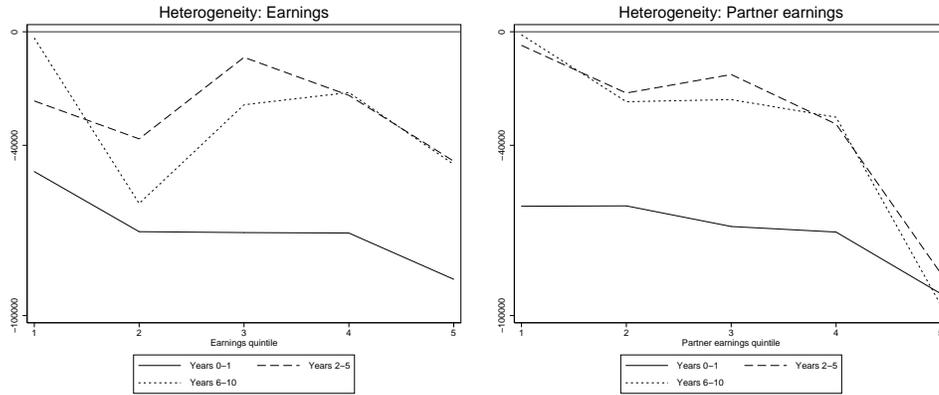
Figure 3: Heterogeneity Analyses

Notes: The figure plots coefficients from regressions on the effect of child birth on annual earnings across the pre-treatment distribution of earning and partner earnings. The pre-treatment characteristics are measured as percentile ranks of the pre-treatment variable under consideration. The coefficients represent regression estimates on samples restricted to income percentiles 1-20, 21-40, 41-60, 61-80 and 81-100. All regressions control for age at IVF treatment, year of treatment, and pre-treatment education and earnings. See text for details.

Table 1: Descriptive Statistics of Selected Variables.

	(1)	(2)	(3)	(4)	(5)
	Non-success	Success	Repr. sample	(2)-(1)	(2)-(3)
Pre-treatment outcomes					
Age at first treatment	32.490	31.415	28.274	-1.075***	3.141***
	(4.445)	(3.886)	(4.297)	(0.069)	(0.060)
Year at first treatment	2000.149	2000.295	2001.446	0.146***	-1.151***
	(3.121)	(3.069)	(4.069)	(0.050)	(0.056)
Annual earnings (1000 DKK)	245.360	243.912	201.717	-1.448	42.195***
	(143.366)	(131.741)	(136.384)	(2.268)	(1.906)
Schooling	12.820	12.843	12.548	0.023	0.295***
	(2.359)	(2.294)	(2.325)	(0.038)	(0.033)
Partner earnings (1000 DKK)	327.006	322.318	287.883	-4.688	34.436***
	(209.665)	(191.939)	(185.995)	(3.464)	(2.722)
Partner schooling	12.678	12.673	12.547	-0.005	0.125***
	(2.389)	(2.323)	(2.316)	(0.040)	(0.034)
Sickness benefits	0.170	0.169	0.143	-0.001	0.026***
	(0.376)	(0.375)	(0.350)	(0.006)	(0.005)
Married	0.521	0.523	0.306	0.002	0.217***
	(0.500)	(0.500)	(0.461)	(0.008)	(0.006)
Positive earnings	0.910	0.922	0.900	0.013***	0.022***
	(0.288)	(0.268)	(0.300)	(0.005)	(0.004)
Full time employment[a]	0.934	0.934	0.842	0.000	0.092***
	(0.248)	(0.249)	(0.365)	(0.004)	(0.005)
Log (hourly wages (DKK))[a]	5.202	5.194	5.084	0.008	0.109
	(0.340)	(0.319)	(0.430)	(0.006)	(0.007)
Weekly hours worked[a]	28.620	28.670	26.691	0.049	1.978***
	(7.734)	(7.729)	(9.083)	(0.137)	(0.137)
Public sector[b]	0.549	0.545	0.444	-0.004	0.101***
	(0.481)	(0.481)	(0.497)	(0.009)	(0.005)
Post-treatment outcomes:					
Annual earnings (1000 DKK)	241.815	211.525	178.907	-30.290***	32.618***
	(144.983)	(128.649)	(127.712)	(2.274)	(1.788)
Positive earnings	0.888	0.864	0.852	-0.024***	0.011**
	(0.267)	(0.282)	(0.349)	(0.004)	(0.005)
Log (hourly wages (DKK))[a]	5.314	5.294	5.255	-0.020***	0.038***
	(0.317)	(0.316)	(0.323)	(0.006)	(0.005)
Hours worked[a]	30.034	27.216	25.390	-2.818***	1.826***
	(26.823)	(21.634)	(7.655)	(0.483)	(0.151)
Observations	13,168	5,370	103,826		

Note: The table shows descriptive statistics for three samples: (1) women having an unsuccessful first IVF treatment, (2) women having a successful first IVF treatment, (3) and a representative sample of Danish women who had their first child born during the study period. Columns (1) to (3) show means with standard deviations in parentheses. Column (4) shows the difference in means between columns 2 and 1 and column (5) shows the corresponding difference in means between columns 2 and 3. Annual earnings are reported in 2008 Danish Kroner (DKK 100 corresponds to USD 20 as of August 2008). [a] Means conditional on having positive earnings. [b] Means conditional on observing sector of work. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table 2: IVF Treatment Effects on Female Labor Outcomes: Results from First-Stage and Reduced-Form Regressions.

Independent variable	Fertility (1)	Earnings (2)	Positive earnings (3)	Weekly hours (4)	Wages (5)	Log wages (6)	Partner earnings (7)	Depression (8)	Divorce (9)
					Panel A: Years 0-1				
IVF success	0.69***	-48,633***	-0.05***	-4.04***	2.90	0.01**	-5,375**	-0.01***	-0.01***
	(0.00)	(1,439)	(0.00)	(0.13)	(2.21)	(0.00)	(2,470)	(0.00)	(0.00)
N	18,538	18,538	18,538	14,022	14,022	14,022	16,689	18,538	18,538
F-statistic	38,427								
					Panel B: Years 2-5				
IVF success	0.32***	-9,402***	-0.01***	0.48***	-8.69***	-0.03***	-3,523	0.00	-0.01***
	(0.00)	(1,703)	(0.00)	(0.11)	(1.44)	(0.01)	(3,004)	(0.00)	(0.00)
N	18,435	18,435	18,435	12,332	12,332	12,332	16,590	18,435	18,435
F-statistic	6,281								
					Panel C: Years 6-10				
IVF success	0.23***	-6,960***	-0.00	0.10	-5.35***	-0.02***	-5,082	0.00	0.00
	(0.00)	(2,397)	(0.00)	(0.13)	(1.86)	(0.01)	(4,536)	(0.00)	(0.00)
N	13,779	13,779	13,779	9,627	9,627	9,627	12,367	13,779	13,779
F-statistic	2,273								
Baseline mean	-	223,038	0.90	28.63	183.01	5.16	301,683	0.05	0.05
Pre-treatment	-	874	0.01***	0.52	-0.06	0.00	-1,298	-0.01	0.00
effect	-	(1,811)	(0.00)	(0.38)	(1.16)	(0.01)	(2,800)	(0.00)	(0.03)

Notes: This table shows first-stage and reduced form regression estimates on the effect of IVF treatment success (0/1) on various outcomes measured at t=0-1, 2-5, and 6-10. In column 1, the coefficient represents the effect of IVF success on the probability of having children during the time period considered. In the reduced form regressions, the coefficient represents the effect of IVF success on the average of the outcome during the time period considered. Time period t=0 refers to the year of the (potential) child birth. All regressions control for age at first IVF treatment, year of first IVF treatment, and pre-treatment average of the outcome studied taken over years 1 to 4 before the first IVF treatment. There are two exceptions. The first-stage regression does not include the pre-treatment average because it is zero by construction. The depression regression includes the pre-treatment value at t-1 because data on antidepressants are only complete from 1994 onwards. The F-statistics in the table refers to F-tests of the significance of the instrument in the first-stage regressions. The baseline mean refers to the mean of the outcome taken over years 1 to 4 before the first IVF treatment (for the sample observed 0-1 years after the year of the (potential) childbirth). The pre-treatment effect refers to the reduced form effect of success at first IVF treatment on the pre-treatment baseline mean. Standard errors are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Fertility Effects on Female Labor Market Outcomes: Results from Instrumental Variable Regressions.

Independent variable	Earnings (1)	Positive earnings (2)	Weekly hours (3)	Wages (4)
		Panel A: Years 0-1		
Fertility	-70,088***	-0.07***	-5.91***	4.24
	(2,054)	(0.01)	(0.19)	(3.23)
% impact	*-31*	*-8*	*-21*	*2*
N	18,538	18,538	14,022	14,022
		Panel B: Years 2-5		
Fertility	-29,378***	-0.04***	1.47***	-26.80***
	(5,285)	(0.01)	(0.35)	(4.45)
% impact	*-12*	*-5*	*5*	*-13*
N	18,435	18,435	12,332	12,332
		Panel C: Years 6-10		
Fertility	-30,675***	-0.02	0.48	-25.18***
	(10,546)	(0.02)	(0.63)	(8.76)
% impact	*-11*	*-2*	*2*	*-11*
N	13,779	13,779	9,627	9,627
Baseline mean	223,038	0.90	28.63	183.01
Pre-treatment	874	0.01***	0.52	-0.06
effect	(1,811)	(0.004)	(0.38)	(1.16)

Notes: This table shows instrumental variable estimates of the effect of having children on various labor market outcomes at t=0-1, 2-5, and 6-10. The coefficients represent the effect of having children. Having children in year t is instrumented with IVF treatment success. Time period t=0 refers to the year of (potential) child birth. All regressions control for age at first IVF treatment, year of first IVF treatment, pre-treatment education, and the pre-treatment average of the outcome studied taken over years 1 to 4 before the first IVF treatment. The baseline mean refers to the average of the outcome taken over years 1 to 4 before the first IVF treatment. The pre-treatment effect refers to the reduced form effect of success at first IVF treatment on pre-treatment baseline mean. Standard errors are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Fertility Effects on Job Changes and Job Characteristics: Results from Instrumental Variable Regressions.

Independent variable	Occ. change (1)	Firm change (2)	Occ. earnings (3)	Occ. (log)wages (4)	Occ. hours (5)	Firm earnings (6)	Firm (log)wages (7)	Firm hours (8)	Firm gender ratio (9)	Public sector (10)	Distance (km) (11)
					Panel A: Years 0-1						
Fertility	0.02*	0.04***	-3,687***	0.00	-0.23***	-1,197	-0.00	-0.05	-0.01**	-0.01	1.38**
	(0.01)	(0.01)	(1,090)	(0.00)	(0.08)	(1,003)	(0.00)	(0.06)	(0.00)	(0.05)	(0.57)
N	17,941	18,194	18,019	18,019	18,019	18,264	18,195	18,194	18,280	17,914	14,440
					Panel B: Years 2-5						
Fertility	0.05	0.04	-4,881*	-0.01	-0.06	-5,761**	-0.02**	0.32	-0.00	0.01	-3.67***
	(0.03)	(0.03)	(2,679)	(0.01)	(0.20)	(2,528)	(0.01)	(0.20)	(0.01)	(0.01)	(1.08)
N	14,166	18,058	17,934	17,934	17,934	18,169	18,072	18,084	18,188	17,824	10,618
					Panel C: Years 6-10						
Fertility	0.01	-0.01	-5,029	-0.01	-0.38	-7,063	-0.02	-0.10	0.00	0.02	-5.53***
	(0.04)	(0.04)	(4,962)	(0.01)	(0.32)	(4,839)	(0.01)	(0.45)	(0.02)	(0.03)	(2.15)
N	11,680	11,782	13,470	13,470	13,470	13,595	13,522	10,974	13,616	13,337	5,173
Baseline mean	0.54	0.57	221,346	5.22	22.33	244,369	5.20	25.58	0.63	0.55	12.58
Pre-treatment effect	-0.005	-0.01	-269	-0.00	0.13*	-1,078	-0.00	0.02	-0.00	-0.00	0.64*
	(0.01)	(0.01)	(990)	(0.00)	(0.07)	(992)	(0.00)	(0.06)	(0.00)	(0.01)	(0.36)

Notes: This table shows instrumental variable estimates of the effect of having children on various labor market outcomes at t=0-1, 2-5, and 6-10. The coefficients represent the effect of having children. Having children in year t is instrumented with IVF treatment success. In columns 1 and 2, the outcome is a binary indicator taking the value of 1 if the woman was recorded as changing occupation or workplace in the short, medium, and long run. Time period t=0 refers to the year of the (potential) child birth. All regressions control for age at first IVF treatment, year of treatment, and pre-treatment education and pre-treatment outcomes in years 1 to 4 before the first IVF treatment. The baseline mean refers to the average of the outcome taken over years 1 to 4 before the first IVF treatment. The pre-treatment effect refers to the reduced form effect of success at first IVF treatment on pre-treatment baseline mean. Standard errors are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Fertility Effects on Female Labor Earnings: Heterogeneity Analyses

Indicator I	Earnings Qrt.4 (1)	Schooling $\geq$15 yrs (2)	Age $\geq$32 yrs (3)	Partner earnings Qrt.4 (4)	Sector Public (5)	Time $\geq$2002 (6)
			Panel A: Fertility effects in reference group sample (I=0)			
0-1 yrs	-64,033***	-68,039***	-63,904***	-64,411***	-76,627***	-66,553***
	(2,215)	(2,300)	(3,019)	(2,343)	(3,353)	(2,542)
2-5 yrs	-20,418***	-26,245***	-16,392*	-15,600***	-36,637***	-30,600***
	(5,383)	(5,479)	(8,890)	(5,896)	(8,726)	(6,483)
6-10 yrs	-19,238*	-28,538***	-16,313	-13,963	-35,523**	-31,324***
	(10,520)	(10,233)	(19,598)	(11,719)	(16,940)	(11,647)
% impact:						
0-1 yrs	*-35*	*-35*	*-33*	*-31*	*-33*	*-30*
2-5 yrs	*-10*	*-12*	*-8*	*-7*	*-15*	*-13*
6-10 yrs	*-8*	*-12*	*-6*	*-5*	*-13*	*-12*
			Panel B: Fertility effects in indicator group sample (I=1)			
0-1 yrs	-86,275***	-74,280***	-74,282***	-89,586***	-67,281***	-75,023***
	(4,705)	(4,294)	(2,772)	(4,804)	(2,638)	(3,410)
2-5 yrs	-48,647***	-36,130***	-36,733***	-79,865***	-24,212***	-28,442***
	(13,956)	(13,274)	(6,513)	(14,666)	(6,817)	(8,909)
6-10 yrs	-60,621**	-31,692	-39,166***	-99,870***	-35,383**	-29,683
	(29,989)	(32,502)	(11,999)	(31,571)	(14,030)	(23,910)
% impact:						
0-1 yrs	*-24*	*-24*	*-30*	*-31*	*-28*	*-32*
2-5 yrs	*-13*	*-12*	*-14*	*-27*	*-10*	*-11*
6-10 yrs	*-16*	*-9*	*-14*	*-31*	*-13*	*-11*
			Panel C: Differences between the two fertility effects			
0-1 yrs	-23,786***	-6,241	-10,377**	-25,175***	9,346**	-8,471**
	(5,286)	(4,872)	(4,107)	(5,345)	(4,268)	(4,254)
2-5 yrs	-33,018**	-9,885	-20,342*	-64,265***	12,425	-2,159
	(15,237)	(5,479)	(11,022)	(15,808)	(11,074)	(11,019)
6-10 yrs	-42,915	-3,153	-22,853	-85,907***	139	-1,640
	(32,255)	(34,075)	(22,980)	(33,676)	(21,996)	(26,597)
Reference mean	182,382	205,626	185,204	205,772	245,156	218,403
Indicator mean	364,445	258,029	249,279	280,355	220,329	229,016

Notes: Panels A and B show instrumental variable regressions on the effect of having a child on earnings at t=0-1, 2-5, and 6-10 on subsamples split by pre-treatment earnings, education, age, partner pre-treatment earnings, sector, and timing of (potential) birth. Year 0 refers to the year of the (potential) child birth. All regressions control for age at first IVF treatment, year of first IVF treatment, and pre-treatment education and earnings. Panel C shows the difference between the subsample instrumental variable estimates. The means reported at the bottom part refer to outcome averages taken over years 1 to 4 before the first IVF treatment in reference group and indicator group samples. Standard errors are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Fertility Effects on Female Labor Earnings: Intensive Margins.

	Extensive margin	Intensive margin		
Instrument	*IVF success*	*IVF success*	*Twins*	*Twins*
Sample	*IVF sample*	*IVF sample*	*IVF sample*	*Representative sample*
	(1)	(2)	(3)	(4)
		Panel A: Years 0-1		
Fertility	-70,088***	-52,686***	-14,507***	-13,052***
	(2,054)	(3,198)	(2,930)	(2,197)
% impact	*-31*	*-24*	*-7*	*-8*
N	18,538	4,598	4,557	103,826
		Panel B: Years 2-5		
Fertility	-29,378***	-9,477*	-6,049	-3,824
	(5,285)	(5,151)	(5,661)	(4,574)
% impact	*-12*	*-4*	*-3*	*-2*
N	18,435	4,581	4,540	103,178
		Panel C: Years 6-10		
Fertility	-30,675***	4,518	-13,154	-2,648
	(10,546)	(7,434)	(9,935)	(9,938)
% impact	*-11*	*2*	*-5*	*-1*
N	13779	3,290	3,543	72,987
Baseline mean	223,038	210,064	221,786	174,156
Pre-treatment	874	6,794*	-4,616	12,073***
effect	(1,811)	(3,816)	(3,592)	(2,767)

Notes: Column 1 replicates the main IV findings of Table 3. Column 2 shows the effect for women with children entering their first IVF treatment. Column 3 shows the effect of family size on earnings for women who had a successful first IVF treatment. Average family size at t=0-1, 2-5, and 6-10 is instrumented with an indicator of having a singleton versus twin birth at first birth. Column 4 shows the corresponding results for a sample of representative, non IVF-treated women. Time period t=0 refers to the year of the (potential) child birth. All regressions control for age at first IVF treatment, year of first IVF treatment, and pre-treatment education and earnings. The baseline mean refers to the average of the outcome variable taken over years 1-4 before the year of the first IVF treatment. The pre-treatment effect refers to the reduced form effect of success at first IVF treatment (columns 1 and 2), or of having a twin birth (columns 3 and 4), on the pre-treatment baseline mean. Standard errors are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendices for Online Publication

Appendix A: Treatment exogeneity

Table 1 contains estimates from regressions linking treatment success (at the first IVF treatment with embryo implant) to pre-treatment labor market outcomes, controlling for year-of-treatment and age-at-first-treatment fixed effects. We find that the estimates attached to the pre-treatment labor market characteristics are all small and, apart from participation, far from statistically significant. The F statistics, reported at the bottom, further indicate that the pre-treatment labor market outcomes together do not predict treatment success.

Table 1: Pre-treatment characteristics and treatment success.

Pre-treatment characteristics	(1) Success	(2) Success
Years of schooling	0.002	0.001
	(0.002)	(0.002)
Annual labor income (DKK)	-0.000	-0.000
	(0.000)	(0.000)
Sickness benefits	-0.010	-0.007
	(0.009)	(0.010)
Marriage	0.002	-0.003
	(0.007)	(0.007)
Positive earnings	0.027**	0.029*
	(0.014)	(0.015)
Partner's years of schooling		0.002
		(0.002)
Partner's annual labor income (DKK)		-0.000
		(0.000)
F-test, pre-treatment labor market characteristics	1.81	1.26
R-squared	0.02	0.02
Observations	18,538	16,689

Notes: The table shows regressions on the probability of success at first IVF treatment. Column 1 controls for age at first treatment, year of first treatment, and various labor market characteristics, measured the year before IVF treatment. Column 2 adds controls for the partner's pre-treatment earnings and years of schooling. The F-test statistic is a test statistic for joint significance of labor market variables. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix B: Some Sensitivity Checks

In Section IV.D, we perform several sensitivity checks to see whether our IVF treatment effect estimates are subject to a number of biases: omitted variable bias related to female health factors, sample selection bias related to wages being observed only for women who work, and sample selection bias related to the unbalanced nature of our short, medium and long-run samples. Full results are reported in Table 2 below.

Table 2: Sensitivity Checks.

Independent variable	Earnings controlling for medical factors	Earnings	Pre-treatment education	Pre-treatment earnings	Earnings
Sample	*Full sample*	*Healthy women*	*Worker sample*	*Worker sample*	*Balanced sample*
	(1)	(2)	(3)	(4)	(5)
	Panel A: Years 0-1				
IVF success	-47,196***	-46,211***	0.01	0.00	-49,114***
	(1,464)	(2,200)	(0.04)	(1.19)	(1,644)
N	18,538	7,378	14,022	14,022	13,779
	Panel B: Years 2-5				
IVF success	-8,088***	-7,379*	0.03	0.36	-10,266***
	(1,728)	(2,667)	(0.04)	(1.94)	(1,954)
N	18,436	7,330	12,332	12,332	13,779
	Panel C: Years 6-10				
IVF success	-5,524**	-7,953**	0.04	2.68	-6,960***
	(2,434)	(3,847)	(0.05)	(2.44)	(2,397)
N	13,779	5,228	9,627	9,627	13,779
Baseline mean	223,042	222,041	-	-	220,508
Pre-treatment effect	1,773	1,270	-	-	1,275
	(1,813)	(2,823)	-	-	(2,055)

Notes: Column 1 reruns the reduced form earnings specification (column 2 in Table 2) adding controls for the following medical factors: number of eggs retrieved, number of embryos implanted, diagnoses, causes of infertility, type of IVF treatment, and clinic indicators. Column 2 shows the effect of IVF treatment success on labor earnings for a sample of women who are *not* diagnosed with any fertility problem. Columns 3 and 4 show the effect of IVF treatment success on pre-treatment schooling and pre-treatment earnings (baseline mean). Column 5 reruns the reduced form earnings specification on a balanced panel of women. Time period $t=0$ refers to the year of the (potential) child birth. All regressions control for age at first IVF treatment, year of first IVF treatment, and pre-treatment education and earnings (except for the regressions in columns 3 and 4). The baseline mean refers to the mean of the outcome variable taken over years 1-4 before the year of the first IVF treatment. The pre-treatment effect refers to the reduced form effect of success at first IVF treatment on the pre-treatment baseline mean (only shown in the relevant regressions). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To be sure that our are wage results are correct and not driven by selective

labor market withdrawal, we have also run our wage regressions on extended samples including non-working women with zero wages. We run reduced-form OLS regressions as well as reduced-form Median Quantile regressions. These results are tabulated in Table 3 and show (not unexpectedly) that the short-run wage responses are very different from the ones we report in the paper. We find that the short-run effect of having children on wages turns significantly negative, which is consistent with the short-run effect we find on labor market withdrawal. In the medium and long run, however, the wage responses are remarkably similar. Because of this similarity, we feel quite certain that medium and long-run wage results are not driven by sample selection.

Table 3: Reduced-Form Wage Regressions using All Women

Independent variable	Wages (wages>0)	Wages (zeros for missings)	Wages (median regression)
	Panel A: Years 0-1		
Success	2.90	-12.28***	-5.29***
	(2.21)	(2.18)	(0.95)
N	14,022	18,538	18,538
	Panel B: Years 2-5		
Success	-8.69***	-5.84***	-4.43***
	(1.43)	(1.80)	(1.25)
N	12,332	18,435	18,435
	Panel C: Years 6-10		
Success	-5.35***	-5.11**	-4.34***
	(1.86)	(2.23)	(1.35)
N	9,627	13,779	13,779

Notes: Column 1 replicates our main (reduced form) wage regression, where inclusion in the sample is conditional on observing positive wages. Column 2 shows the reduced form effect of IVF success on wages where missing wages take the value 0. Column 3 shows median regression results (with zeros for missing). Time period $t=0$ refers to the year of the (potential) child birth. All regressions control for age at first IVF treatment, year of first IVF treatment, pre-treatment education, and average wages taken over 1 to 4 years before IVF treatment. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix C: Delayed Fertility of Always Takers

Most IVF treated women end up having children despite a first failed IVF attempt. We call this delayed fertility. If the impact of having children on female labor earnings is larger when children are young, delayed fertility will affect our fertility effect estimates. To see how this works, we introduce a simplified three-period version of our fertility model and two groups of women with different demands for children (or different command over resources). Let

women from group a exert a weaker demand for children; these are women who would remain childless after a first failed IVF attempt. Such women always earn Y_a in the absence of children. After a successful first treatment, these women experience an earnings loss C_a in period 1 when children are young, and c_a in periods 2 and 3 when children are older. Let women from group b exert a stronger demand for children; these are women who would end up having children in period 2, despite a first failed IVF attempt. Such women always earn Y_b in the absence of children. After a successful first treatment, they experience an earnings loss C_b in period 1, and c_b in periods 2 and 3 when children are older. After a failed first treatment, such women have a child in period 2 and experience an earnings loss C_b in period 2 when children are young, and c_b in period 3 when children are older. Let β_a and β_b represent the shares of women from group a and b who receive a successful first treatment. In the context of IV estimation, these two groups represent compliers and always takers (Imbens and Angrist 1994). If we compare fertility and earnings differences between successfully and unsuccessfully treated women in period 1, we obtain the first-stage and reduced-form effects in a simplified fertility model without covariates

$$E(F_1|Z{=}1) - E(F_1|Z{=}0) = \beta_a + \beta_b,$$

$$E(Y_1|Z{=}1) - E(Y_1|Z{=}0) = -\beta_a C_a - \beta_b C_b.$$

Combining these two equations gives us the IVF-based IV estimator

$$\lambda_1^{IV} = -\frac{\beta_a}{\beta_a + \beta_b}C_a - \frac{\beta_b}{\beta_a + \beta_b}C_b = -C_a + \frac{\beta_b}{\beta_a + \beta_b}[C_a - C_b]$$

The IV estimator for period 1 captures the effect of having any children on labor earnings, which is the weighted average of the causal effects in groups a and b. In a similar spirit, we can construct the IV estimator for periods 2 and 3. Combining the fertility and earnings differences in periods 2 and 3 between women who were successfully and unsuccessfully treated in period 1, we get the IVF-based IV estimators

$$\lambda_2^{IV} = -c_a + \frac{\beta_b}{\beta_a}[C_b - c_b],$$

$$\lambda_3^{IV} = -c_a.$$

The IV estimator for period 2 captures the causal effect of interest and a nuisance parameter, which is the delayed fertility effect of women in group b who experienced a failed treatment in period 1. Note that the delayed

fertility effect will reduce the earnings of these women relative to those who were successful in period 1, so the estimated effect of interest is biased towards zero. The IV estimator in period 3 solves this problem.

More generally, the fertility effect of interest represents a mixture of fertility and delayed fertility effects on labor earnings. While delayed fertility is a concern, it will give us conservative fertility estimates that are still informative about the impact children have on their mothers' labor earnings.

Appendix D: All IVF and Twin Samples

In this paper we use four different samples in estimation: (i) a sample of IVF treated women without children (baseline sample); (ii) a sample of IVF treated women with children; (iii) a sample of successfully treated IVF women (with at least two embryo implants for the twin strategy); and (iv) a sample of representative women. Table 4 provides sample means and standard deviations for a selected number of variables.

Appendix E: Other External Validity Results

Another way to investigate the external validity of our IVF results is to compare labor earnings of women in our primary IVF sample to those of women in a more representative sample drawn from the full population of women who had their first born child around the same time as IVF treated women had their first IVF attempt. Figures 1a and 1b show simple descriptive graphs for average labor earnings for these two groups of mothers, six years preceding the year of childbirth, for the year of childbirth, and for ten years following childbirth. We plot two measures of labor earnings: conditional and unconditional labor earnings. The conditional earnings measure refers to the year-to-year dummy parameters in regressions that control for pre-treatment education and age at having the first child (Figure 1b). The unconditional earnings measure refers to year-to-year average earnings, regardless of the individual characteristics (Figure 1a). In terms of unconditional earnings, one can see that before as well as after childbirth the labor earnings of IVF treated women are consistently higher than those of the representative sample. In terms of conditional earnings, one can see that the decline in earnings is somewhat sharper among the IVF women and the increase in earnings is somewhat steeper in the representative sample of women. Nevertheless, both measures of labor earnings in the IVF and representative samples follow roughly the same pattern before

44

Table 4: Descriptive Statistics of Selected Variables from All Samples.

	IVF sample (main sample)	IVF sample (previous children)	IVF sample (twin trick)	Representative sample
	(1)	(2)	(3)	(4)
Pre-treatment outcomes:				
Age at first IVF treatment	32.18	34.97	31.52	28.27
	(4.32)	(4.08)	(3.91)	(4.30)
Year at first IVF treatment	2000.20	2000.50	2000.00	2001.45
	(3.11)	(3.11)	(2.99)	(4.07)
Annual earnings (1000 DKK)	244.9	226.1	244.3	218.3
	(140.1)	(150.7)	(131.0)	(132.9)
Schooling	12.83	12.50	12.81	12.55
	(2.34)	(2.47)	(2.30)	(2.32)
Sickness benefits	0.17	0.19	0.17	0.14
	(0.37)	(0.39)	(0.38)	(0.35)
Married	0.52	0.65	0.53	0.31
	(0.50)	(0.48)	(0.500)	(0.46)
Positive earnings	0.91	0.87	0.92	0.90
	(0.28)	(0.33)	(0.27)	(0.30)
Post-treatment outcomes:				
Annual earnings (1000 DK)	233.04	254.69	234.48	178.91
	(141.11)	(138.39)	(117.95)	(127.71)
Positive earnings	0.88	0.94	0.94	0.85
	(0.27)	(0.16)	(0.16)	(0.35)
Observations	18,538	4,598	4,557	103,826

Notes: This table shows descriptive statistics for the following samples and subsamples. Column (1): childless women entering their first IVF treatment. Column (2): women with children entering their first IVF treatment. Column (3): women with a successful first IVF treatment who had at least one child during the study period. Column (4): a representative sample of women who had at least two embryos inserted. Column (4): a representative sample of women who had at least one child during the study period. Standard deviations within parentheses. Annual earnings are reported in 2008 Danish Kroner (DKK 100 corresponds to USD 20 as of August 2008).

and after childbirth, which is suggestive that any inherent differences between these two groups of mothers are probably not the leading cause of any differences in how they respond to first born children.
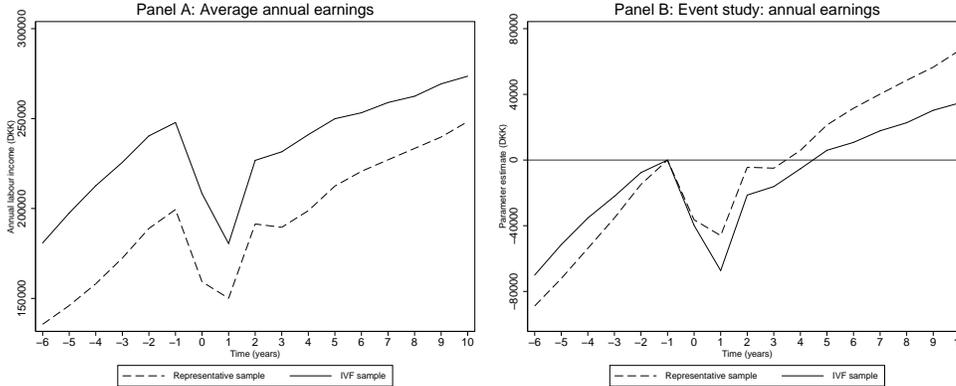

Figure 1: Relationships between (First) Child Birth and Annual Earnings.

Notes: The figures represent unconditional and conditional labor earnings. The unconditional earnings measure refers to year-to-year averages (Figure 1a). The conditional earnings measure refers to coefficients from (before and after) OLS regressions on the relationship between child birth and annual earnings controlling for the year of child birth and age of having a first child (Figure 1b). We have normalized coefficients to 0 in the year before child birth. We use the sample of IVF treated women and the sample of representative women who had their first child between 1995-2005. See text for details.

As an additional check, we run naive OLS regressions of annual labor earnings on having children, age and year fixed effects and years of education using the two samples of women (ignoring that fertility is endogenous). Results are reported in Table 5. If any inherent differences between the women with and without IVF treatment would lead to different labor supply responses to childbearing, we expect to find different associations between annual labor earnings and fertility, measured at the extensive margin. This is not what we observe. The fertility estimates in the IVF treated sample of women, which are all negative and statistically significant in the short, medium, and long run, do not differ much from those found in the sample of representative women. Also the time series patterns of the fertility estimates are roughly the same. We therefore believe that unobserved differences between women with and without IVF treatment are an unlikely reason for differences in labor supply responses.

Table 5: External Validity Check: Results from Naive Least Squares Regressions of Annual Labor Earnings on Having Children.

Independent variable	Earnings	Earnings
Sample	*IVF sample*	*Representative sample*
	(1)	(2)
	Panel A: Years 0-1	
Fertility	-54,176***	-51,226***
	(1,252)	(472)
N	184,976	529,017
	Panel B: Years 2-5	
Fertility	-18,475***	-23,431***
	(1351)	(498)
N	184,354	528,428
	Panel C: Years 6-10	
Fertility	-7,260***	-12,157***
	(1,647)	(565)
N	170,563	524,511

Notes: Column 1 shows OLS (panel) estimates of the relationship between having a first child and annual earnings for the sample of IVF treated women. Column 2 shows the corresponding estimates for the sample of representative women. Time period $t=0$ refers to the year of the (potential) child birth. All regressions control for age at first child birth, year of child birth, and education. Standard errors are clustered at the individual level and are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix F: Other US Twin Studies

We have surveyed three US studies on the impact of children on female labor market outcomes using twins at first birth as an instrumental variable. We take annual labor earnings as the main labor market outcome, focus on impact estimates taken from the 1980 (or more recent) Census Public Use Micro Samples (PUMS), and convert each estimate into percentage point changes in average annual earnings to facilitate the comparison. Bronars and Grogger (1994) study the effect of having twins at first birth on the labor earnings of unwed mothers with children under 18 taken from the 1980 PUMS. On the full sample, they find that earnings decline 15 percentage points for having a second child. Using a sample of mothers with children between 0 to 3 years old, they find that earnings decline 22 percentage points. Bronars and Grogger also report estimates when children are older, but only the short-run estimate is large enough to be statistically significant. Jacobsen, Pearce and Rosenbloom (1999) also examine the effect of having twins at first birth on the labor earnings of all mothers (including unwed mothers) with children under 18 in the 1980 PUMS. They report overall impacts as well as short, medium and long-

run impacts. They find an overall decline in average annual labor earnings of 5 percentage points. When they take the children's age into account, they find that annual labor earnings decline 22 percentage points when children are 0 to 2 years old, 8 percentage points when children are 3 to 5 years old, and 4 percentage points when children are 6 to 10 years old. Again, only their short-run estimate is statistically significantly different from zero. Since they report only the reduced-form impact of having twins on female labor earnings, we have rescaled their estimates with the short, medium and long-run first-stage estimates (which are approximately 0.96, 0.67 and 0.60). Vere (2011) uses twins at first birth to examine the effect of having two or more children on the labor earnings of all women who are between 21 and 35 years at the time of sampling in the 1980, 1990 and 2000 PUMS. He finds that having an additional second child reduces earnings by 38, 24 and 20 percentage points using the 1980, 1990 and 2000 PUMS respectively. Because the children in these samples are younger, on average, Vere finds fertility effect estimates that are larger than the ones reported in the other two studies.

When we use twins at first birth as instrument for having additional children on a representative sample of all Danish women, we find that having additional children reduces earnings, on average, by 3 percentage points, regardless of the age of the children (not reported). When we take the children's age into account, we find that having additional children reduces earnings by 8 percentage points in the short run, 2 percentage points in the medium run, and 1 percentage point in the long run. Also we find statistically significant negative fertility effects in the short run, but not in the medium and long run. Overall, we find that all the fertility effect estimates reported in these US studies are larger than those obtained in Denmark.