Fertility Effects on Female Labor Supply: IV Evidence from IVF Treatments*

Petter Lundborg, Erik Plug and Astrid Würtz Rasmussen

October 2014

Abstract

This paper introduces a new IV strategy based on IVF induced fertility variation in childless families to estimate the causal effect of having children on female labor supply using 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 are: (*a*) negative, large and long lasting; (*b*) much stronger at the extensive margin than at the intensive margin; and (*c*) similar for mothers, not treated with IVF, which suggests that IVF findings have a wider generalizability.

In this paper we examine how children affect their mothers' labor supply. In particular, we focus on the extensive fertility margin and estimate the labor supply response to having children among women who have no children (yet). To identify fertility effects at the extensive margin, we propose a novel instrumental variables strategy based on in vitro fertilization (IVF), in which we treat IVF treatment success at the first IVF treatment as a natural experiment. 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 to have children among childless women and can be used to estimate the causal effect of having any children on female labor supply.

The IVF strategy further provides useful insights into the intensive fertility margin. First, we can estimate the labor supply response to having additional children

*Erik Plug (corresponding author): Amsterdam School of Economics, University of Amsterdam, The Netherlands, Tinbergen Institute, IZA and UCLS (email: e.j.s.plug@uva.nl). Petter Lundborg: Department of Economics, Lund University, IZA (email: petter.lundborg@nek.lu.se). Astrid Würtz Rasmussen: Department of Economics and Business, Aarhus University (email: awr@econ.au.dk). We thank seminar and conference participants in Aarhus, Amsterdam, Berlin, Braga, Edinburgh, Essen, Göteborg, Hamburg, Helsinki, Lund, Mannheim, Stockholm, Tilburg, Uppsala and York for their comments and suggestions. Data in this article are, in part, taken from IVF registers, held by the Danish National Board of Health. DNBH is not responsible for the analysis or interpretation of the data presented. Financial support from the Institute for Evaluation of Labour Market and Education Policy (IFAU) and Swedish Research Council for Health, Working Life and Welfare (FORTE) is gratefully acknowledged.

among IVF treated women who already have children and compare the effects of fertility on labor supply on both intensive and extensive margins. Second, we can use the larger fraction of twin births among IVF births and apply the more commonly used IV strategy based on having twins at first birth on a sample of IVF treated women; that is, we compare the labor supply response of IVF treated women having twins with those having singletons. Examples of the twin strategy to estimate the labor supply consequences of having children include Mark Rosenzweig and Kenneth Wolpin (1980), Stephen Bronars and Jeff Grogger (1994), Jaisri Gangadharan et al. (1996) and Joshua Angrist and William Evans (1998). By comparing the twin estimates with those obtained on a more representative population sample, we also offer a potentially valuable way to examine the wider generalizability of IVF findings.

The data we use are a combination of 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 the Danish tax registers which hold records on annual earnings (including parental leave and sickness benefits), hourly earnings and worker absenteeism for a 20 year period running from 1991 to 2011. At Statistics Denmark these registers are matched and complemented with standard demographic variables, such as education, age, gender, marital status and total number of children.

The main finding is that IVF treated women earn persistently less because of children. In particular, we find that (*a*) labor supply responses to having any children are negative, large and long lasting; (*b*) labor supply responses are much stronger at the extensive fertility margin than at the intensive fertility margin; and (*c*) labor supply responses to having twins (as opposed to having singletons) are similar for IVF mothers and other mothers. We therefore believe that the IVF findings we present in this paper are of general interest.

The remainder of the paper proceeds as follows. Section I provides the background and motivation behind this study. Section II discusses IVF treatments and IVF data used in estimation. Section III introduces our empirical strategy. Section IV lists the identifying assumptions and presents some evidence concerning their validity. Section V reports our main set of results. Section VI explores whether the estimated fertility effects hold up against alternative interpretations. Section VII concludes.

I Literature background

Much empirical work on the labor market consequences of childbearing has shown, for years, that women with children fare generally worse than women without children: they work fewer hours, earn lower salaries, have less prestigious jobs, and are less likely to work for pay (Martin Browning 1992; Jane Waldfogel 1998; Roland Fryer and Bruce Sacerdote 2008). When asked about how children affect their mothers' labor supply, however, economists often struggle to give causal answers. This is because most studies cannot separate causation from adverse selection. Adverse selection occurs when women who chose to have children have weaker labor force attachment or lower earnings potential than those who chose to remain childless and may give rise to a spurious interpretation of the observed associations between fertility and labor supply: women with children work and earn less, regardless of having children.

In order to separate causation from adverse selection, 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 female labor supply: twins at first birth and sex composition of the first two children (in combination with parental preferences for mixed sex siblings). Those empirical studies that treat the arrival of twins as a natural 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 (Rosenzweig and Wolpin 1980; Bronars and Grogger 1994; and Gangadharan et al. 1996). Also those empirical studies that treat sex composition of the first two children as a natural experiment find that mothers of same sex children work less than mothers of mixed sex children because of third born children. The same mothers, however, seem to catch up in the longer run (Angrist and Evans 1998; Maria Iacovou 2001; Guillermo Cruces and Sebastian Galiani 2007; Eric Maurin and Julie Moshion 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. The reason is that these experiments create exogenous variation in the number of children, but only among women that 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 therefore ask is how fertility affects labor supply at the extensive margin, where childless women decide to have children or not. Or, in other words, what are the labor market consequences of (involuntary) childlessness? The answer to this question is -we think- important for two reasons. First, labor supply responses measured at the extensive fertility margin are of general interest because they capture the fertility margin to which all mothers are (or have ever been) exposed to. Second, fertility effects are possibly larger when estimated at the extensive fertility margin because simple theories on household production, child quality and economies of scale predict they do.[1]

Unfortunately, we know very little about the labor market consequences of childlessness. Empirical evidence is scarce and mixed. In recent years, there have been a few attempts to analyze the relationship between fertility at the extensive margin and female labor supply (Jorge Agüero and Mindy Marks 2008 2011; Julian Cristia 2008). These studies employ instrumental variable strategies to account for endogenous fertility, but use different instrumental variables. Agüero and Marks (2011), for example, treat self-assessed infertility as a natural experiment and compare fertility rates and labor supply responses between infertile and fertile women aged 20

[1]This view is further shaped by personal experiences. The authors are all parents of two children; they all experienced that their first born child had a much bigger impact on their economic lives than their second born child.

to 44 in 26 developing countries. While the authors observe that self-assessed infertility indeed corresponds to fewer children, their main finding is that fertile women work as much as infertile women, regardless the fertility margin.[2] 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 advise and fertility tests, but exclude in vitro fertilization treatments, which plays a key role in our study. With treatment success taken as a natural experiment, he then compares childbearing and labor supply decisions of childless women who sought help to achieve pregnancy 21 months earlier. His main finding is that women with babies work significantly less than those without.[3]

We believe there is scope for more empirical work on having children and labor supply. The current literature on the extensive fertility margin is in its infancy with many unresolved but promising issues. In this paper, it is our motivation to explore some of these unresolved issues; that is, we apply a novel instrumental variable strategy using a very large sample of IVF treated families to measure short and long run labor market consequences of children for women at both extensive and intensive fertility margins.

II IVF register, treatment and sample

We use administrative register data on all IVF treatments performed in public and private fertility clinics and hospitals in Denmark. In particular, our data set contains information on all women residing in Denmark who received at least one IVF treatment somewhere between 1994 and 2005. To show how IVF data can help us to estimate the causal effect of fertility on labor supply, we start this section by providing more details on the IVF register, IVF treatment, IVF instrument and the primary IVF sample we use in estimation.

A 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 has been mandatory and completeness of IVF registrations is close to a 100 percent. The IVF register holds records on the reason for infertility, mode of treatment, the number of eggs retrieved from the womb, the number of fertilized eggs transferred back, outcomes of treatment (birth, abortion, stillbirth, or not pregnant), date of treatment and date of birth when birth took place. The IVF register has been merged to other administrative registers to get longitudinal information running from 1991 to 2011 on standard demographic and labor market variables, such as education, age,

[2] Agüero and Marks (2008) apply the same instrumental variable strategy based on self-assessed infertility, but on a sample of 6 developing countries in Latin America. They find no impact of motherhood on labor supply, measured as work for pay in the week prior to the survey.

[3] Daniel Aaronson, Fabian Lange, Bhashkar Mazumder (2014) describe an interesting multigenerational experiment on fertility decisions along extensive and intensive margins. A large school construction program starting in the early 1920s directly affected the family lives of two generations of women. Women who faced improved schooling opportunities for their children were more likely to become mother but of fewer children. The children who experienced improved school outcomes were less likely to become mother, but those who did were mothers of fewer children.

gender, marital status, number of children, labor market attachment, the number of hours worked, and annual earnings. The IVF register contains information on 31,767 women receiving all together 96,807 IVF treatments.[4]

B IVF treatment

While IVF treatment is the leading medical intervention to help infertile women to get 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 advise and fertility testing. After a year of having frequent and unprotected sex without getting pregnant, these women are medically diagnosed infertile and can then be referred to a fertility clinic or hospital if they are below the age limit of 40. General practitioners are also responsible for referrals. With a referral, women are entitled to have three IVF treatments on the Danish National Health Care System. Without a referral, women have to pay themselves.[5]

Once referred to a fertility clinic or hospital, women undergo the 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 selection of the most promising embryos. The selected embryos are transferred back into the womb. IVF treatments often fail with much uncertainty at each stage of the treatment: fertility medication does not work, there are no eggs to retrieve, there are no suitable embryos to transfer, or the transferred embryos simply stop growing. The average success rate of treatment is about 22 percent.

Because success rates are quite low, most women undergo multiple treatments to achieve success. But not all. After each unsuccessful treatment, there are women who decide to end treatment. Moreover, the more treatments women undergo, the higher the share of women who decide to end treatment.[6] With IVF treatments being costly in social, psychological as well as financial terms, these rising shares suggest that the process of IVF treatments is selective and that women who decide to continue treatment are probably women with more resources or a stronger demand for children.

C IVF instrument

In order to exploit the IVF process to arrive at the causal link running from fertility to work effort, we need some exogenous shock in the fertility treatment. We know

[4] Of the original 32,073 women in the IVF register, we have removed 306 women for whom labor market information is not available at the year of IVF treatment.

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

[6] According to the IVF register, we observe that 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.

that most women undergo multiple IVF treatments to achieve success. With the number of IVF treatments being endogenous, however, it does not make much sense to treat the success rate in a sequence of IVF treatments as exogenous. Instead, we argue that success in the first IVF treatment may generate exogenous variation in fertility at the extensive margin. In particular, we consider the first IVF treatment where women have successfully reached the fourth stage and got embryo implantation. For some women, the embryos develop and IVF treatment will lead to pregnancy and children. For other women, the embryos stop growing and IVF treatment will fail. If the development of implanted embryos in the womb is to a large extent exogenously determined, exogenous treatment success more or less 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 chances of success are somehow related to the pre-treatment labor supply characteristics we observe in our data.

D IVF sample

From the IVF register, we draw our primary sample of 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 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 multiple treated women are possibly different from firstly treated women in ways related to preference for children or financial resources. Since few women skip a full year between two subsequent treatments, we start counting IVF treatments for those women 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 all the original 31,767 women who were treated at least once somewhere between 1994 and 2005, we remove women who were treated in 1994, women who enter the first IVF treatment with children, and women without suitable embryos at the first IVF treatment. This leaves us with an IVF sample of 18,538 treated women.

Table 1 provides sample means and standard deviations for some of the outcome variables (post-treatment outcomes) we study below and for some of the control variables (pre-treatment characteristics) we use to measure the extent to which successfully and unsuccessfully treated women (and partners) are comparable. The same table also provides sample means and standard deviations for a 30 percent representative sample of women born around the same time as IVF treated women for comparison purposes. The sample consists of 105,922 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.

Two observations follow from this table. First, 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 statistically significant, suggest that IVF treated women are on average women with a stronger earnings potential than other women with children. Second, 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. For example, they attain the same level of education and have exactly the same annual earnings before they seek IVF treatment. After the first treatment, however, the same women turn out to be different; that is, successfully treated women have more children, 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 and work effort, we need a more sophisticated analysis to appropriately identify the impact of fertility on labor supply.

III IV methodology using IVF treatments

To identify fertility effects at the extensive margin, we make use of a standard IV strategy in which success at the first IVF treatment among childless women with embryo implants serves as our main instrumental variable; that is, we estimate a linear relationship between labor supply and fertility of woman i who were first treated t years earlier using a two-stage least squares model where the first stage is

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

and the second stage is

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

In these two equations, Y is a measure of female labor supply, 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 supply, 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 may contain unobservable factors that are either related to fertility, work effort or both. The parameter of interest is δ_t, which measures the effect of fertility on labor supply measured at the extensive fertility margin t years after the first IVF treatment.

There can be some uncertainty about what the treatment year actually is because of the time it takes for an IVF treatment to be successful. Since our focus is on changes in fertility, we define the treatment year as the year of (potential) child birth after a successful (failed) first treatment and estimate these two equations for the ten subsequent years following the first treatment year. In our IVF sample, the median duration of gestation takes nine months for first successful treatments. We therefore calculate the year of potential child birth by adding nine months to the day of embryo transfers and assume that women with a failed treatment would have given birth nine months after the first treatment had their IVF treatment with embryo implants been successful.

IV Is the IVF instrument a valid instrument?

For this IV methodology to work, the IVF instrument must satisfy three conditions, being (a) treatment success is correlated with fertility at the extensive margin (first stage relevance); (b) treatment success is uncorrelated with the econometric error u (second stage independence); and (c) treatment success has no impact on labor supply other than through its first stage impact on fertility (second stage exclusion). Here, we provide graphical evidence in support of instrument relevance and independence. In Section VI, we turn to the exclusion condition and derive some implicit tests to argue its plausibility.

A First stage relevance

Does IVF treatment success at first IVF treatment with embryo implants generate meaningful variation in fertility, measured at the extensive margin? Given that IVF treatment is the leading medical intervention to help infertile women to get pregnant and conceive children, the answer is likely affirmative. Figure 1 provides graphical evidence demonstrating that success is indeed strongly correlated with fertility. In there, we visualize the first stage specification (without control variables) and plot for all successfully treated women in our IVF sample the likelihood to have children for six years preceding the year of giving birth, the year these women have their first child (which we refer to as year zero), and for the ten years following childbirth. In the same figure, we also plot the likelihood to have children for all unsuccessfully treated IVF women before and after year zero, which is the year of potential child birth as defined above. Since we construct our IVF sample to consists of childless women up to IVF treatment entry, we do not see any fertility differences between the two groups of women. In year zero fertility patterns begin to diverge. For successfully treated IVF women we see, almost by definition, that fertility jumps sharply from zero in the year of childbirth. For unsuccessfully treated women we also see that average fertility increases following the year of potential childbirth, despite a first unsuccessful first treatment. This is because most women continue treatment and may be successfully treated in subsequent IVF treatments (or succeed without fertility treatment). In the long run the overall success rate is about 77 percent among women whose first treatment failed. The differences in fertility between women whose first IVF treatment was successful or not, which represents the first stage effect of the IVF instrument on fertility, is always 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 long run. In the results section we will report estimates of the impact of treatment success on fertility measured at the extensive margin.

B Second stage independence

Are IVF treatment success and pre-treatment labor supply truly independent? While we think it is plausible to assume that IVF treatment success at the first treatment with embryo implants contains random elements, the extent to which treatment success is randomly determined is still an unresolved issue. The concern we have is that women whose first IVF treatment lead to a pregnancy and childbirth may have an inherently weaker labor force attachment or lower earnings potential, producing a spurious result that IVF women with children work and earn less. In addition to the observed similarity in pre-treatment labor market outcomes for the two groups of women, which is already suggestive of exogenous treatment success, we provide two extra partial checks. Both checks assert that under treatment success independence we should not see any differences in pre-treatment labor market characteristics.

The first check graphically plots the reduced form specification (without control variables) and shows average labor earnings profiles of women who were either successfully or not successfully treated before and after the first treatment. Figure 2 depicts the annual earnings, for all the IVF treated women in our sample, for six years preceding the year of (potential) childbirth, for the year of (potential) childbirth, and for ten years following (potential) childbirth. Before (potential) childbirth, we see that the average earnings profiles before treatment run virtually identical for the two groups of women. At the time of childbirth, however, we see that successfully treated women experience an immediate and large decline in annual earnings, which we attribute to having children, and not to something else. After (potential) childbirth, we see that differences in earnings between successfully and unsuccessfully treated women are the largest during the first two years, start to decline thereafter but are still there when the first born children are about ten years old. The persistent fall in earnings observed among successfully treated women, accompanied with the persistent higher chance of having any children, is consistent with the view that women work persistently less because of childbearing.

The second check regresses treatment success at the first IVF treatment with embryo implant on a set of pre-treatment labor market outcomes, including measures of average labor earnings, labor market participation, full time work and worker sickness absenteeism. We also include year-of-treatment and age-at-first-treatment fixed effects. Age at first treatment should account for the widely held medical notion that age is considered the single most important factor in assisted reproduction (Zev Rosenwaks, Owen Davis and Mark Damario 1995; Laura van Loendersloot et al. 2010). Table 2 reports these estimates. In column (1) we find that the estimates attached to the labor market outcomes are all very small and far from statistical significant. Also the F statistics, reported at the bottom of Table 2, indicate that the pre-treatment labor market outcomes together do not predict treatment success. In column (2) we find that these estimates do not change when we add several medical characteristics typical to the IVF treatment, including measures on medical indication, type of treatment, number of eggs collected, number of embryo implants, and type of clinic.[7]

In all, our initial concern that women with a successful first IVF treatment may be inherently different from women with a failed first IVF treatment in ways related to the labor supply decisions they make seems to be misplaced. Our results rather suggest that treatment success at first IVF treatment with embryo implants is as good as randomly assigned.

[7] We should note that the dummy variables for the women's age at first treatment are jointly statistically significant. The corresponding F statistics are much larger than those reported for the pre-treatment labor market and indicate that age-at-treatment is a predictive factor in treatment success.

V Main results

We now turn to our main regression results, based on the empirical specification described above. Table 3 contains first stage, reduced form, and instrumental variables estimates for women who underwent their first IVF treatment somewhere between 1995 and 2005 and who had no previous children. In these regressions, we take annual labor earnings as our main outcome measure of labor supply, 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.

The first stage estimates, which we report in panel A, mirror the pattern sketched in Figure 1, where variation in success at first IVF treatment led to a long lasting difference in the chance of having children. The estimates reveal that these fertility differences are statistically significant. At the bottom row of panel A, we learn that success at first IVF treatment is a strong instrument for fertility measured at the extensive margin, with F statistics far beyond the typical rules of thumb values. Recall that these first stage estimates are interpreted as the effect of a successful first IVF treatment on the likelihood that having children or not, evaluated t years after the first treatment, where year zero (t=0) denotes the (potential) birth of a child. In the year of (potential) child birth, for instance, our estimate suggests that a successful first IVF treatment raises the probability of having a child with 81 percentage points. We do not find a 100 percentage increase because most women after a first failed treatment continue treatment the same year, of which some are successful. In subsequent years our fertility estimates fall, but stabilize from the seventh year onwards; that is, the long run likelihood of having children among those women that ever undergo IVF treatment converges to 23 percentage points, meaning that the majority of women who were unsuccessful at the first IVF treatment attempt manage to have children as time passes by.

The reduced form estimates, which we report in panel B, measure the direct effect of a successful first IVF treatment on annual labor earnings 0-10 years after the (potential) birth of the child. Annual labor earnings are reported in Danish Kroner (DKK 100 corresponds to USD 17 as of September 2014). The first thing to note is that all estimates are negative and statistically significant in both the short, medium, and long run. In the short run, we find that the estimated effects are the largest the year of (potential) childbirth and the year thereafter, where annual earnings fall with about DKK 43,000 to DKK 53,000. In the longer run, these negative fertility effects are still there, albeit smaller. Ten years after the (potential) birth of a child, we find that the effect of a successful first IVF treatment reduces earnings by DKK 10,000. When interpreting these estimates, it is important to keep in mind that the reduced form effect of DKK 10,000 is driven by those 23 percent IVF women with children. To arrive at the causal impact of having children or not, we need to scale up our reduced form effects by the corresponding first stage effects. This is what we do next.

The instrumental variables estimates, which we report in panel C, indicate that the labor supply response is greatest in the short run, where our estimates imply that having children reduces annual earnings with DKK 51,000 to DKK 98,000 during the two first years. The larger effect observed the year after the birth of the child can be explained in two ways. First, mothers have been working for a larger or lesser part during the year of child birth, depending on the timing of birth. In addition, almost all mothers take up maternity leave and receive their full salary, with a cap, during the first 6 months of maternity leave. Second, most mothers take up maternity leave for an extended period at a lower compensation rate. This means that for some mothers, the full impact of maternity leave take up materializes the year following childbirth. As we move further ahead in time, the IV estimates decline in magnitude, but remain statistically significant. Three years after the (potential) childbirth, having children reduces annual earnings by DKK 44,000 and 6 years after by DKK 36,000. Even 10 years after the (potential) childbirth, we find that earnings are DKK 47,000 lower because of children. Since average pre-treatment earnings amount to about DKK 230,000, these effect estimates suggest that women earn roughly 15 to 20 percent less because of children. These estimates indicate that the long run labor market consequences of childbearing, measured at the extensive margin, are substantial.

A Other female labor market outcomes

Female annual labor earnings, which is our main labor supply outcome, depends on the number of hours women work in a year as well as the hourly wage rate. To further explore where the observed fall in annual labor earnings comes from, we analyze the impact of having children on female labor force participation, working full-time, wage rates and job changes.[8] Table 4 contains IV estimates for these alternative labor market outcomes. In panel A we report the IV estimates for annual labor earnings for ease of comparison. In panels B to F we report IV estimates for the labor outcomes labor force participation, full-time work, hourly wages and a job change indicator. In this table we run similar IV regressions, except that we replace outcome and pre-treatment controls for the alternative labor market measure under study.

We first consider the number of hours women work. If the fall in annual earnings is driven by women working less, we should find that women with children are more likely to either stop working or replace full-time for part-time work. In panel B we show the labor market participation response to having children. In the shorter run, we find that labor force participation rates are significantly lower due to childbearing. These fertility effects are most notable during the first two years of motherhood, when participation rates fall with 6 to 7 percent points. In the longer run, however, we find that labor force participation rates are not affected by having children. The longer run fertility effects we estimate are all statistically insignificant and very close to zero. In Panel C we show whether women work fewer hours and replace full-time for part-time work because of children. These labor supply effects do not differ in any material way from those reported in panel B, with women working fewer hours when children are young. It seems that women tend to stay at home or work fewer hours when first born children are young, but

[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 the alternative labor market outcomes for all workers. This implies that regression results for working full-time, hourly earnings and log hourly earnings are obtained on smaller samples. In addition, there is uncertainty about the accuracy of the hourly wage measure of part-time workers.

return to the labor market when children turn three or four and then continue to participate as much and work as many hours as women without children. As such, these results cannot explain the large medium and long run impacts of children on earnings we observed above.

We next consider the wage rate. If the fall in annual earnings is not driven by women working less, it must mean that women get lower hourly wages, at least in the longer run. The results for hourly wages, reported in panels D and E, confirm this. 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 long run. To put these numbers into context, we report the estimates for log hourly earnings and find that hourly wages fall with 9 to 16 percent in the medium and long run.

We can speculate about possible explanations. Because of children, women may have less work experience, lag behind in wage negotiations, get fewer promotions, or choose to work in more family friendly but lower paid jobs. To get some insights on the latter explanation, we check whether differences in hourly earnings may come from women changing occupations. Our data allow us to classify occupations into one-digit occupational groups. If the occupation in year t differs from the occupation held the year before treatment, we can infer that this is a new occupation. When we estimate the impact of having children on the likelihood of changing occupations, which we report in panel F, we do not find much.

B Extensive versus intensive fertility margins

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 their parents' labor supply is larger at the extensive margin than at the intensive margin, just as simple theories predict.

In light of this, IVF treatments prove (again) 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 instrument on a sample of IVF treated women who already have children when they start their treatment. The instrument IVF treatment success 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, and going from two to three or more children for mothers who enter treatment with two children. We should note that the vast majority of these mothers are no longer eligible for free treatments and pay themselves. 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 instrument on a sample of successfully treated women who received at least two embryo implants. In this case, we zoom in on women 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 instrument to raise the likelihood of going from one to two (or more) children.

Table 5 contains IV estimates of fertility effects on labor supply, measured at both extensive and intensive margin. In panels A and B we report IV fertility estimates at the extensive margin, similar to those in Tables 3 and 4, for two different samples: a sample of childless women entering their first IVF treatment (which is our baseline sample) and a sample of childless women entering their fourth IVF treatment.[9] In comparison to women in our baseline sample, women at the fourth IVF attempt must pay for their treatment and thus, on average, express a stronger demand for children. As expected, we find that the fertility estimates for women at the fourth IVF attempt are mostly larger than the corresponding fertility estimates for women at the first IVF attempt. These estimates provide further evidence that the short, medium and long run labor market consequences of childbearing, measured at the extensive margin, are substantial.

In the next two panels we report IV fertility estimates at the intensive margin. In panel C we show fertility estimates that come from IV estimation using the first IVF treatment as instrument on a sample IVF treated women with children. In comparison to the previous two samples, these women share either treatment order (panel A) or willingness to pay (panel B). In the short run, we find substantial and significant labor supply responses, albeit somewhat smaller in size than the labor responses observed at the extensive margin. In the medium run, however, we find that the fertility estimates get much smaller and, in most cases, lose their statistical significance. In the long run, we find that some of the estimates even turn positive, although being less precisely estimated. In panel D we show the same fertility estimates but use twins at first birth as instrument on a sample IVF treated women with twins and singletons. In the short, medium and long run, we find relatively weak labor supply responses to changes in family size. We find significant estimates in the two years following twin birth, but those are much smaller in size than the ones we obtained at both extensive and intensive margins reported in previous columns. This is because most mothers take up full maternity leave after childbearing, regardless of giving birth to twins. In the medium and long run, we find that fertility estimates are small, insignificant and a bit jumpy, being sometimes positive and sometimes negative. These medium and long run estimates are comparable to those reported in the previous panel, but also to those reported elsewhere in the literature which exploit fertility variation due to twinning and sibling sex composition. 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.

[9] Similar to the fertility variation at the first IVF attempt, we can also exploit fertility variation at the fourth IVF attempt. Conditional on treatment failure at the first three treatment, those women entering the fourth treatment are again very similar, apart that for some women the fourth treatment was successful. In this case, we can again apply instrumental variable estimation using the fourth IVF treatment as instrument to estimate fertility effects at the extensive margin for childless women entering the fourth treatment. This allows us to test whether the effect of childlessness is comparable across different populations of IVF treated women (panels A and B). In addition, we should note that most of these women pay for the IVF treatment themselves, as the Danish healthcare system provides three free treatments to infertile childless couples. This further allows us to compare fertility effects along extensive and intensive margins among IVF treated women who share their willingness to pay for treatment (panels B and C).

C 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 supply 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 investigate this is to compare labor earnings responses to first born children (extensive fertility margin) of women in our 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. Figure 3 present 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. One can see that before and after childbirth labor earnings of IVF treated women are consistently higher than those of the representative sample. Nevertheless, the labor earnings 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 their labor supply responds to first born children.[10]

Another way of examining 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 supply responses. The natural experiment we have in mind is the twin experiment, which generates fertility variation at the intensive margin. We conjecture that with comparable labor supply responses, IVF results are generalizable. In Table 5, panels D and E, we report fertility estimates based on the twin experiment for two samples: a sample of successfully treated IVF women and a sample of representative women. We find that the fertility estimates are remarkably similar, particularly in the shorter run. For instance, one additional child the year following birth leads to a fall in earnings of DKK 25,000 for IVF treated mothers and of DKK 23,000 for non-treated IVF mothers. In the year after, the fall in earnings equals to DKK 16,000 and DKK 19,000, respectively. We also find that fertility estimates are statistically similar in the longer run, but we readily confess that the labor supply differences are rather imprecisely estimated.

In sum, we find that IVF treated women and other women, not treated with IVF, respond in similar ways to both endogenous fertility shocks, measured at the extensive margin, and exogenous fertility shocks, measured at the intensive margin. If observed similarity in labor supply responses is informative about unobserved

[10] As an additional check, we ignore that fertility is endogenous and 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. Results are reported in Appendix C. If any inherent differences between the women between women with and without IVF 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 are an unlikely reason for differences in labor supply response.

similarity in labor supply responses to exogenous fertility shocks, measured at the extensive margin, our IVF findings generalize to a larger population of representative women.

VI Alternative interpretations

Our IVF estimates suggest that the effect of having children on labor supply is negative, substantial and long lasting. Some uncertainty about the causal interpretation, however, may arise if the exclusion restriction fails and IVF treatment success affects labor supply in some other way than through the increased likelihood of having children. Of particular concern is that we may wrongfully attribute the observed fall in earnings to having children. To provide some evidence on the validity of the exclusion restriction, we introduce three incriminating mechanisms in which childbearing is not the main reason why women work less after childbearing and test for their presence in our data.

A Divorce

The first mechanism is a divorce mechanism. Women who divorce (or anticipate divorce) generally work longer hours (see, for example, Kelly Bedard and Olivier Deschenes 2005; Betsey Stevenson and Justin Wolfers 2007; and Olivier Bargain et al. 2012). If a failed treatment stirs up marital instability, we may find that unsuccessfully treated women work (and earn) more than successfully treated women because of higher divorce risks and not because of reduced fertility. Figure 4, in which we plot marriage rates for all (un)successfully treated women at first treatment for six years preceding and ten years following (potential) childbirth, does not show such a divorce pattern. Women who seek IVF treatment face similar divorce risks, regardless of treatment success.[11]

B Opportunity costs of working

The second mechanism is an opportunity cost mechanism. After an unsuccessful treatment, women may reassess their labor market situation, experience reduced opportunity cost of working, and decide to work longer hours, apply for promotions and more lucrative jobs (or women are disappointed and decide to work more as a coping strategy). To see whether this is an issue, we zoom in on childless women who decide to end IVF treatment after a failed first IVF treatment. Because these women express a weaker demand for children than those who decide to continue IVF treatment, they arguably form an opportunity cost sensitive control group for which opportunity cost driven labor responses should materialize more clearly through accelerated earnings growth after the first failed treatment. Figure 5, in which

[11] While the production of children is widely recognized as the most important source of marital gains (Gary Becker 1981, Yoram Weiss 1997), the effect of children on marital stability has proved very difficult to estimate for well-known selection reasons. In an accompanying paper, we apply a comparable IVF strategy to analyze how children affect marital stability in much more detail (Petter Lundborg, Erik Plug and Astrid Würz Rasmussen 2015). In there, we find that children do not improve long run marital stability and conclude that children do not contribute much to long run marital specific capital.

we examine the earnings profiles of women without children who ended treatment after a failed first IVF treatment, goes against women working more after a failed treatment. Instead, we find that these women tend to work somewhat less, shortly after the first failed treatment. Perhaps IVF treatments carry some health risks that not only restrict the short run labor supply but also make some women decide to terminate treatment. While this is a concern, it will give us conservative fertility estimates that are still informative about the impact children have on their mothers' labor supply.[12]

C Delayed fertility

The third mechanism is a fertility delay mechanism. Most IVF treated women end up having children despite a first failed IVF attempt. In comparison to women with children after the first IVF treatment, those women who are successful in subsequent IVF treatments must wait a year or longer to have children. We call this delayed fertility, with possible implications for the long run labor supply estimates using IVF treatment success at first treatment as instrument. If fertility effects on labor supply are larger when children are younger, our concern is that the long run labor market consequences of having children we observe in our data may come from short run labor market consequences of women with delayed success in second or higher IVF treatments. To deal with this issue, we explore differences in shares between women who decide to end treatment after a failed first and fourth treatment (6 percent versus 30 percent). If long run fertility effects are driven by short run fertility effects of women who delayed fertility, we should see that the larger share of women who terminate treatment after a fourth consecutive failure lead to weaker long run fertility effects among childless women entering the fourth treatment. This is not the case. In panels A and B of Table 5 we find that long run fertility effects for women at the fourth IVF attempt are mostly larger, not smaller, than the corresponding fertility estimates for women at the first IVF attempt.

A related concern, one that it shares with most other fertility studies in this field, is that delayed fertility may have a direct influence on female labor supply. If so, the exclusion assumption would again not hold. To provide some basic insights on the role of delayed fertility in IV fertility estimates based on IVF treatments, we present a simplified version of the fertility model we described earlier. Suppose there are two groups of women a and b with different demands for children (or different command over resources) among the childless women who enter their first IVF treatment. Women from group a exert a weaker demand of children; these are women who would remain childless after a first failed IVF attempt. Women from group b exert a stronger demand of children; these are women who would end

[12] In Appendix D we address this health concern in more detail. In there, we present results of three additional regressions that are (somehow) informative about unhealthy women, treatment failure, and poor labor market performance: (i) we reestimate our baseline IV regression adding rich medical information (possibly reflecting health problems) gathered at the time of treatment, including the number of eggs collected and transferred, diagnoses, causes of infertility, type of IVF treatment, and clinic indicators; (ii) we reestimate our baseline IV regression on a subsample of (arguably) healthy women, i.e. where the fertility problem is on the partner's side; and (iii) we reestimate our reduced form regression using the intake of antidepressants as outcome variable. In both IV regressions, fertility results remain the same and stay significant. In the reduced form regression, we find no relationship between a failed treatment and depression. We all take this as evidence that health risks (possibly induced by IVF treatments) are not our biggest concern.

up having children anyway, despite a first failed IVF attempt.[13] If β_t and $1-\beta_t$ represent the shares of women from group a and b, we know that fertility differences between successfully and unsuccessfully treated women t years after first treatment can be written as a weighted sum of the fertility differences in group a and b

$$E(F_t|Z=1) - E(F_t|Z=0) =$$
$$\beta_t(E(F_t|Z_a=1) - E(F_t|Z_a=0)) + (1-\beta_t)(E(F_t|Z_b=1) - E(F_t|Z_b=0)),$$

which is the first stage effect in a simplified fertility model without covariates. If women in group b with a failed first IVF attempt have had their children, the first stage effect equals β_t. This follows by construction: after an unsuccessful IVF treatment women in group a remain childless t years after first treatment ($E(F_t|Z_a=1) - E(F_t|Z_a=0) = 1$), whereas all the women in group b end up having children ($E(F_t|Z_b=1) - E(F_t|Z_b=0) = 0$). In a similar spirit, we can write the labor supply differences between successfully and unsuccessfully treated women as a weighted sum of the labor supply differences in group a and b; that is,

$$E(Y_t|Z=1) - E(Y_t|Z=0) =$$
$$\beta_t(E(Y_t|Z_a=1) - E(Y_t|Z_a=0)) + (1-\beta_t)(E(Y_t|Z_b=1) - E(Y_t|Z_b=0)).$$

The IV estimate of δ in this simplified setup equals

$$(3) \qquad \delta_t^{IV} = \underbrace{E(Y_t|Z_a=1) - E(Y_t|Z_a=0)}_{\delta_{at}} + \frac{1-\beta_t}{\beta_t}\underbrace{E(Y_t|Z_b=1) - E(Y_t|Z_b=0)}_{\delta_{bt}}$$

where the first part measures the effect of fertility on labor supply at the extensive margin among women with a weaker demand for children δ_{at}, and the second part measures the effect of delayed fertility on labor supply among women with a stronger demand for children δ_{bt}. The fertility effect we estimate is therefore a mixture of fertility and timing on labor supply, which depends on signs and magnitudes of known and unknown parameters β_t, δ_{at} and δ_{bt}. Distinguishing between fertility effects δ_{at} and timing effects δ_{bt} is important.

Equation (3) is a linear equation with two unknown parameters from which we can back out the average values for δ_{at} and δ_{bt}. Because we have yearly measures of δ_t^{IV} and β_t covering a ten year period, we can simply regress δ_t^{IV} on $(1-\beta_t)/\beta_t$ and interpret the estimates of intercept and slope parameters as average values of δ_a and δ_b, respectively. If we do this, we find point estimates for δ_a and δ_b of -64.214 [4.24] and 10.546 [1.84], respectively (with t-values between brackets). The combination of a much larger fertility effect with a barely significant but positive timing effect seems to suggest that the high IV estimates we find are primarily driven by fertility and not by delayed fertility.[14]

[13] In the context of IV estimation, these two groups represent the compliers and always takers (Imbens and Angrist 1994). Although the long run fertility at the extensive margin of always takers is not affected by the instrument, always takers do contribute to the IV estimate because of possible labor market consequences of delayed fertility.

[14] Previous studies on delayed fertility in which age at menarche and miscarriage are used to generate exogenous variation in the timing of birth often find that delayed fertility has a positive but modest effect on the labor market outcomes of mostly young mothers (Arnaud Chevalier and Tarja Viitanen 2003; Joseph Hotz, Susan Williams McElroy and Seth Sanders 2005; Amalia Miller 2011). Our positive but barely significant estimates of delayed fertility on annual labor earnings of mostly older mothers appear similar.

It is obvious that none of these findings suggest that childbearing is the only causal channel through which treatment success affects female labor supply. Nonetheless, all these findings are supportive of the widely held view that labor market consequences of childbearing are substantial and long lasting.

VII Concluding remarks

This paper evaluates how fertility choices made at extensive and intensive margins affect labor supply 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 at 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 some of the questions we have posed earlier. First, do women earn less because of first born children (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. 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 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 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 supply of IVF treated women in Denmark, where parents have access to generous maternity leave, job protection and child care arrangements. In particular, questions arise about the wider generalizability of IVF findings. First, are the labor supply responses 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 other women do. Second, how do women respond to children in countries that have more restrictive maternity leave arrangements? With our data at hand, an answer is not readily available. But we can speculate about possible answers. 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. On the other hand, we may find larger fertility effects if women who would otherwise return to their employer at the end of the maternity leave period leave the labor market altogether. We view this as one of the exciting empirical questions for future research.

References

Aaronson, Daniel, Fabian Lange, and Bhashkar Mazumder, "Fertility Transitions along the Extensive and Intensive Margins," *American Economic Review*, 2014, *104* (11), 3701–24.

Agüero, Jorge and Mindy Marks, "Motherhood and Female Labor Force Participation: Evidence from Infertility Shocks," *American Economic Review*, May 2008, *98* (2), 500–504.

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.

Ahn, Namkee and Pedro Mira, "A Note on the Changing Relationship between Fertility and Female Employment Rates in Developed Countries," *Journal of Population Economics*, 2002, *15* (4), 667–682.

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).

Boca, Daniela Del, Silvia Pasqua, and Chiara Pronzato, "Motherhood and Market Work Decisions in Institutional Context: A European Perspective," *Oxford Economic Papers*, 2009, *61* (1), 147–171.

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.

Chevalier, Arnaud and Tarja Viitanen, "The Long-Run Labour Market Consequences of Teenage Motherhood in Britain," *Journal of Population Economics*, 2003, *16* (2), 323–343.

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.

Gangadharan, Jaisri, Joshua Rosenbloom, Joyce Jacobson, and James Wishart Pearre, "The Effects of Child-Bearing on Married Women's Labor Supply and Earnings: Using Twin Births as a Natural Experiment," NBER Working Papers 5647, National Bureau of Economic Research, Inc July 1996.

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.

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.

Hotz, Joseph, Susan Williams McElroy, and Seth Sanders, "Teenage Childbearing and its Life Cycle Consequences Exploiting a Natural Experiment," *Journal of Human Resources*, 2005, *40* (3), 683–715.

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.

Lundborg, Petter, Erik Plug, and Astrid Würtz Rasmussen, "Children and Divorce," Working Paper, University of Amsterdam 2015.

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.

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.

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

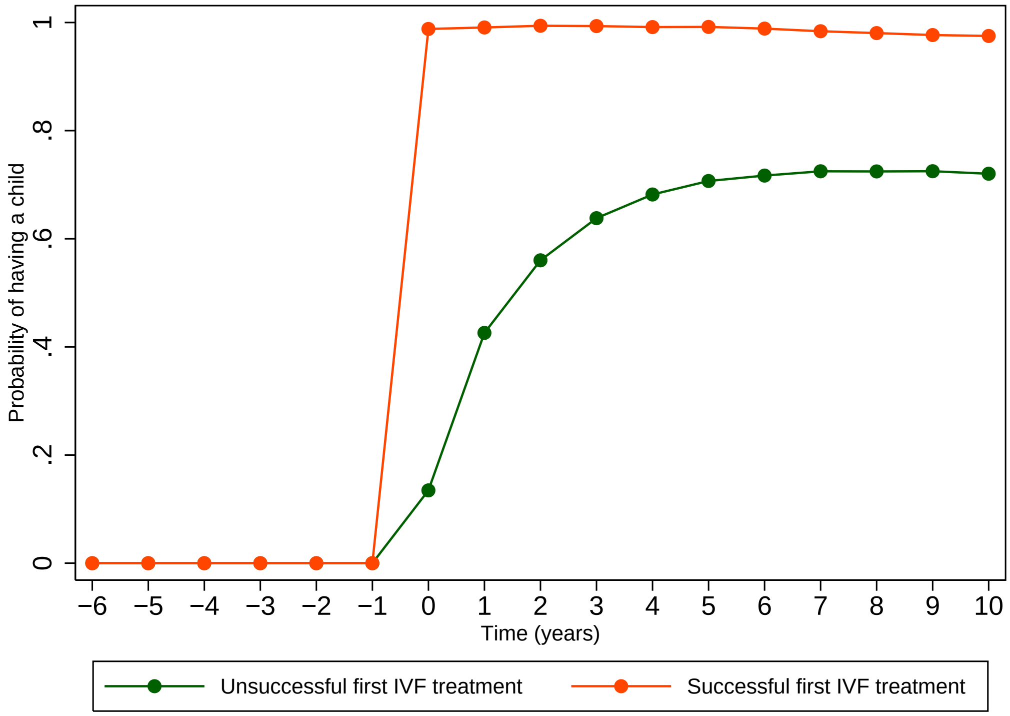
Templeton, Allan, Joan Morris, and William Parslow, "Factors that Affect Outcome of In-Vitro Fertilisation Treatment," *The Lancet*, 1996, *348* (9039), 1402–1406.

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.

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

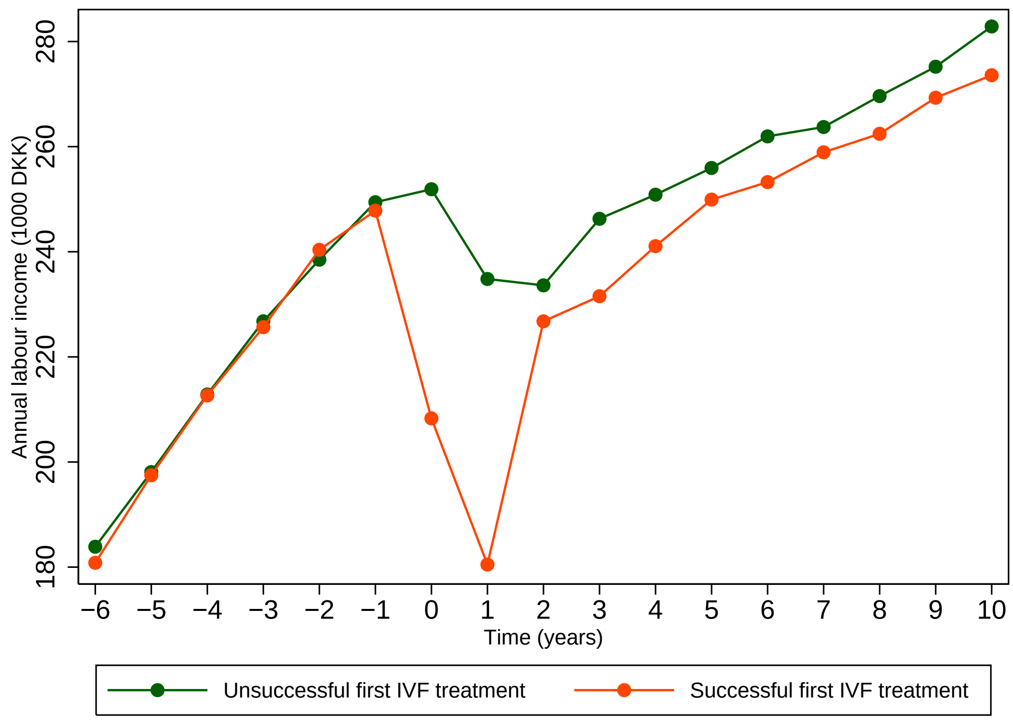
Weiss, Yoram, "The Formation and Dissolution of Families: Why Marry? Who Marries Whom? And What Happens upon Divorce," *Handbook of Population and Family Economics*, 1997, *1*, 81–123.

Figure 1: Fertility at the extensive margin before and after the first IVF treatment.



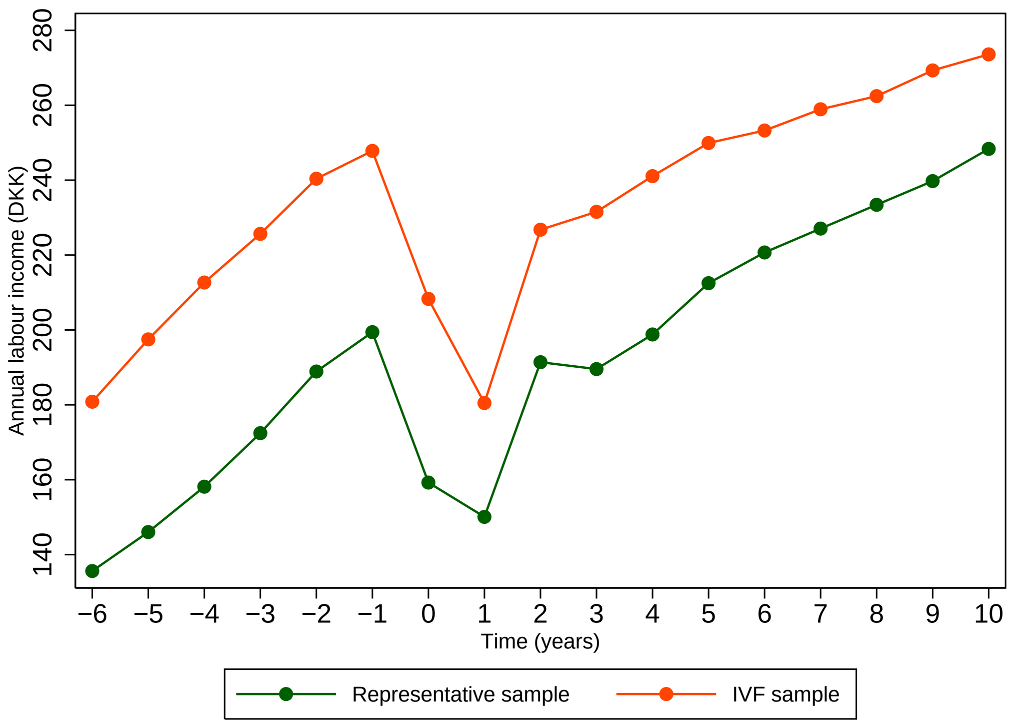
Note: The figure plots the probability of having a child by IVF treatment success. Sample consists of childless women entering their first IVF treatment 1995-2005. See text for details. Year zero denotes the year of the (potential) child birth.

Figure 2: Annual earnings before and after the (potential) birth of a child.



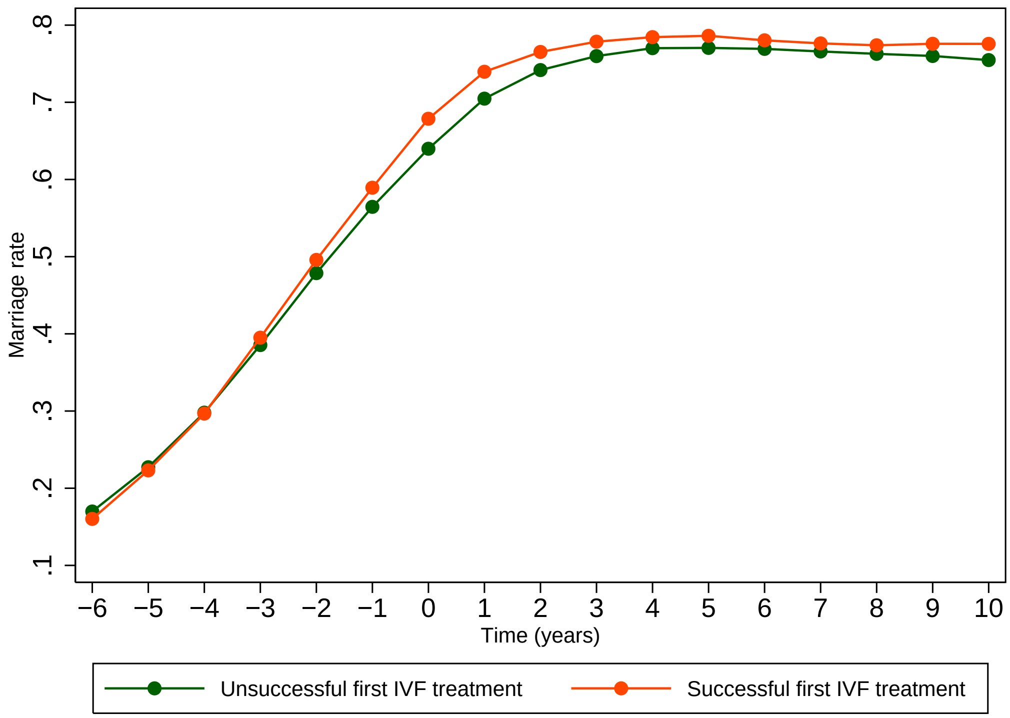
Note: The figure plots mean annual labor income in Danish krones (DKK) by IVF treatment success. Sample consist of childless women entering their first IVF treatment 1995-2005. See text for details. Year zero denotes the year of the (potential) child birth.

Figure 3: Annual earnings before and after the birth of a first child. IVF sample and representative sample.



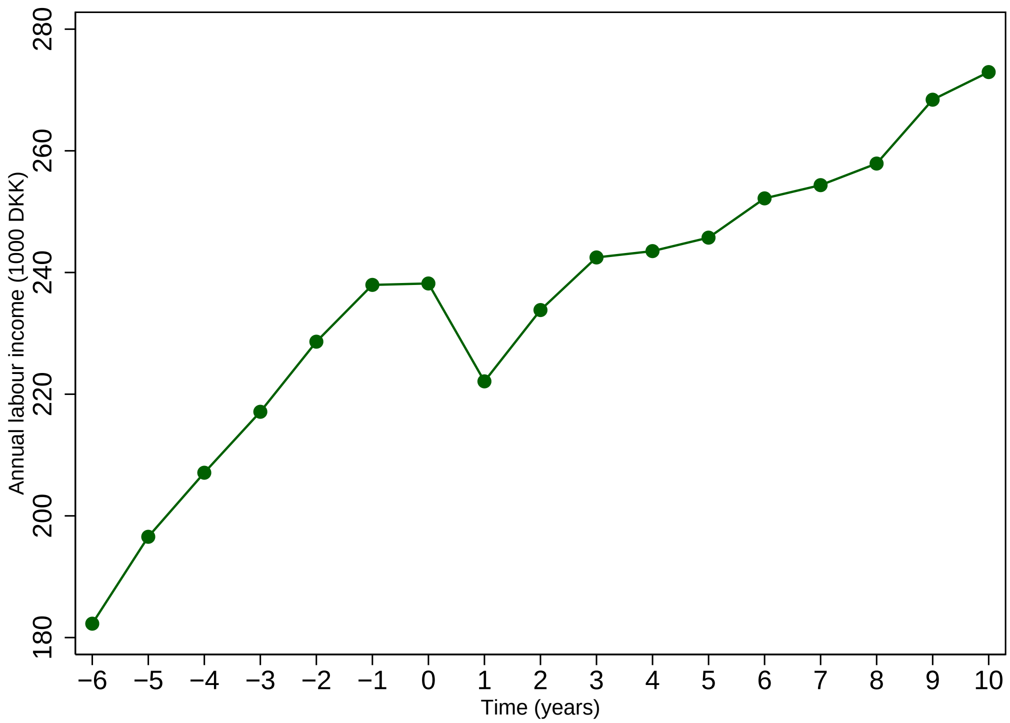
Note: The figure plots mean annual labor income in Danish krones (DKK) before and after the birth of the first child for (1) childless women entering their first IVF treatment between 1995-2005 and (2) a representative sample of women who had their first birth between 1995-2005. Year zero denotes the birth of the first child.

Figure 4: Marriage rates before and after the (potential) birth of a child.



Note: Marriage rates before and after the (potential) birth of a child. The figure plots marriage rates by IVF treatment success. Sample consist of childless women entering their first IVF treatment 1995-2005. See text for details. Year zero denotes the year of the (potential) child birth.

Figure 5: Annual earnings before and after the (potential) birth of a child. Sample of women who quit after an unsuccesful first IVF treatment.



Note: The figure plots mean annual labor income in Danish krones (DKK) for women entering their first IVF treatment between 1995-2005 and who quit after a unsuccesful first IVF treatment. Year zero denotes the year of the (potential) child birth.

Table 1: Descriptive statistics.

	(1)	(2)	(3)
	Non-success	Success	Representative sample
Age at first treatment	32.49	31.41	28.30
	(4.445)	(3.886)	(4.282)
Year at first treatment	2000.1	2000.3	2001.3
	(3.121)	(3.069)	(4.008)
Pre-treatment labor income in 1000s DKK (years 1-3)	231.5	229.2	184.7
	(132.4)	(121.8)	(121.8)
Pre-treatment years of schooling	12.82	12.84	12.56
	(2.359)	(2.294)	(2.320)
Pre-treatment incidence of sickness benefits	0.170	0.169	0.143
	(0.376)	(0.375)	(0.350)
Married year before treatment	0.521	0.523	0.308
	(0.500)	(0.500)	(0.462)
Pre-treatment labour market participation	0.909	0.922	0.902
	(0.288)	(0.268)	(0.297)
Pre-treatment full time employment	0.862	0.874	0.780
	(0.345)	(0.332)	(0.414)
Post-treatment labor income in 1000s DKK (years 0-3)	241.8	211.5	181.1
	(145.0)	(128.6)	(127.3)
Post-treatment labor market participation (years 0-3)	0.888	0.864	0.826
	(0.267)	(0.282)	(0.306)
Post-treatment full time employment (years 0-3)	0.841	0.815	0.731
	(0.299)	(0.305)	(0.362)
Observations	13169	5369	105922

Note: The table shows descriptive statistics for three samples: (1) sample of childless women entering their first IVF treatment and failing the attempt, (2) corresponding sample of women succeeding in their first IVF attempt, (3) representative sample of Danish women who had their first child during the study period.

Table 2: Pre-treatment characteristics and treatment success.

Variables	(1) Success	(2) Success
Pre-treatment years of schooling	0.001	0.002
	(0.002)	(0.002)
Pre-treatment labor income in 1000s DKK (years 1-3)	-0.000	-0.000
	(0.000)	(0.000)
Incidence of sickness benefits year before treatment)	-0.010	-0.013
	(0.009)	(0.009)
Married year before treatment	0.002	0.003
	(0.007)	(0.007)
Pre-treatment labor market participation	0.016	0.010
	(0.016)	(0.016)
Pre-treatment full time employment	0.021	0.011
	(0.014)	(0.014)
Medical controls		✓
F-test age-at-treatment	69.25	11.60
F-test pre-treatment labor market characteristics	1.518	1.494
R-squared	0.019	0.049
Observations	18538	18538

Notes: The table shows regressions on the probability of success at first IVF treatment. Column 1 controls for age at first treatment. Column 2 controls age at first treatment and for the following medical factors: the number of eggs collected, the number of embryo implants, type of IVF treatment, medical indications for infertility, and clinic indicators. F-test statistics are test statistics for joint significance of age-at-tretament and labor market variables, respectively. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Fertility effects on female labor earnings: Results from first-stage, reduced form, and instrumental variable regressions.

Independent variable	(1) t=0	(2) t=1	(3) t=2	(4) t=3	(5) t=4	(6) t=5	(7) t=6	(8) t=7	(9) t=8	(10) t=9	(11) t=10
Panel A: First stage regressions using any children (0/1) as dependent variable											
IVF success (0/1)	0.84***	0.54***	0.41***	0.33***	0.28***	0.26***	0.24***	0.23***	0.23***	0.23***	0.23***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
N	18538	18494	18435	18381	17404	15599	13779	11983	10173	8342	6620
F-stat.	57424	13303	7844	5493	4055	3199	2482	1874	1520	1176	943
Panel B: Reduced form regressions using annual earnings as dependent variable											
IVF success (0/1)	-43299***	-53082***	-6009***	-14340***	-9050***	-5269**	-8679***	-5992**	-7536**	-5566	-10975***
	(1541)	(1700)	(1830)	(1934)	(2055)	(2244)	(2463)	(2738)	(3038)	(3491)	(3981)
N	18538	18494	18435	18381	17404	15599	13779	11983	10173	8342	6620
Panel C: IV regressions using annual earnings as dependent variable											
Any children (0/1)	-51251***	-97914***	-14781***	-43798***	-32147***	-20493**	-35866***	-26064**	-32814**	-24338	-47079***
	(1826)	(3099)	(4474)	(5872)	(7275)	(8710)	(10168)	(11901)	(13237)	(15262)	(17114)
N	18538	18494	18435	18381	17404	15599	13779	11983	10173	8342	6620

Notes: This table shows first-stage regressions, reduced form regressions, and instrumental variables regressions. Panel A shows the effect of a successful first IVF treatment on the probability of having any children betwen t and t=10. Panel B shows the effect of a successful first IVF treatment on earnings between t and t+10. The columns show estimates from separate regressions. Time period t=0 refers to the year of the (potential) child birth. All regressions control for age at IVF treatment, year of treatment, and pre-treatment education and earnings. Standard errors are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Fertility effects on other female labor market outcomes: Results from instrumental variable regressions.

Independent variable	(1) t=0	(2) t=1	(3) t=2	(4) t=3	(5) t=4	(6) t=5	(7) t=6	(8) t=7	(9) t=8	(10) t=9	(11) t=10
	Panel A: Annual earnings (baseline)										
Any children (0/1)	-51251***	-97914***	-14781***	-43798***	-32147***	-20493**	-35866***	-26064**	-32814**	-24338	-47079***
	(1826)	(3099)	(4474)	(5872)	(7275)	(8710)	(10168)	(11901)	(13237)	(15262)	(17114)
N	18538	18494	18435	18381	17404	15599	13779	11983	10173	8342	6620
	Panel B: Labor force participation (0/1)										
Any children (0/1)	-0.06***	-0.08***	-0.04***	-0.05***	-0.02	-0.03	-0.03	-0.01	-0.03	-0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)
N	18538	18494	18435	18381	17404	15599	13779	11983	10173	8342	6620
	Panel C: Full-time work (0/1)										
Any children (0/1)	-0.02***	-0.02***	-0.03**	-0.03**	-0.05***	-0.02	-0.04	-0.05	0.00	-0.03	0.02
	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)
N	16414	16067	16184	16273	15484	13897	12188	10582	8985	7385	5863
	Panel D: Hourly wages										
Any children (0/1)	6.88**	-0.52	-16.86***	-14.85***	-20.57***	-23.51***	-23.47***	-25.28***	-34.46***	-20.84*	-37.99**
	(3.42)	(3.50)	(3.61)	(4.79)	(5.99)	(7.38)	(8.42)	(8.51)	(9.32)	(12.28)	(18.05)
N	13733	12985	12512	11288	10025	8689	7372	6024	4764	3680	2557
	Panel E: Log hourly wages										
Any children (0/1)	0.02**	-0.00	-0.08***	-0.08***	-0.09***	-0.09***	-0.12***	-0.12***	-0.16***	-0.12**	-0.13**
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)	(0.07)
N	13733	12985	12512	11288	10025	8689	7372	6024	4764	3680	2557
	Panel F: Job change (0/1)										
Any children (0/1)	-0.04***	-0.06***	-0.03*	-0.02	-0.00	-0.02	-0.06	-0.02	-0.08*	-0.04	0.02
	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)	(0.06)
N	18510	18468	18409	18358	17382	15580	13763	11968	10160	8333	6614

Notes: Panel A replicates the main IV findings of Table 3. Panel B shows the effect of having any children on labor market participation between t and t+10. Panels C to F show the corresponding results for working full time, the wage rate, an an indicator of job change. Time period t=0 refers to the year of the (potential) child birth. The models in panels A and E control for age at IVF treatment, year of treatment, pre-treatment education and earnings. In B to D pre-treatment earnings are replaced with the relevant pre-treatment outcome, i.e. labor market participation, full time work, and hourly wage rate, respectively. Time period t=0 refers to the year of the (potential) child birth. Standard errors are in parentheses;* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Fertility effects on female labor earnings: Various results for extensive and intensive fertility margins

Independent variable	(1) *t*=0	(2) *t*=1	(3) *t*=2	(4) *t*=3	(5) *t*=4	(6) *t*=5	(7) *t*=6	(8) *t*=7	(9) *t*=8	(10) *t*=9	(11) *t*=10
	Panel A: Earnings results for childless women entering first IVF treatment (extensive fertility margin)										
Any children (0/1)	-51251***	-97914***	-14781***	-43798***	-32147***	-20493**	-35866***	-26064**	-32814**	-24338	-47079***
	(1826)	(3099)	(4474)	(5872)	(7275)	(8710)	(10168)	(11901)	(13237)	(15262)	(17114)
N	18538	18494	18435	18381	17404	15599	13779	11983	10173	8342	6620
	Panel B: Earnings results for childless women entering fourth IVF treatment (extensive fertility margin)										
Any children (0/1)	-47920***	-111731***	-39870***	-56509***	-50886***	-50316***	-33615**	-22410	-46379**	-29464	-15136
	(3906)	(5359)	(7068)	(9358)	(11138)	(13072)	(15022)	(17688)	(19789)	(22122)	(25564)
N	4299	4291	4276	4263	4028	3603	3165	2664	2256	1852	1477
	Panel C: Earnings results for women with children entering first IVF treatment (intensive fertility margin)										
Number of children	-39739***	-73100***	-4860	-9158	-13854**	-11887*	-4530	6109	15340*	10572	8770
	(2929)	(4460)	(5049)	(5668)	(6426)	(6841)	(7417)	(8802)	(9095)	(10426)	(12311)
N	4598	4590	4581	4566	4278	3754	3290	2791	2306	1885	1488
	Panel D:Earnings results for IVF treated mothers with singletons and twins (intensive fertility margin)										
Number of children	-3881	-24678***	-15781***	-2061	2471	-8125	-12607	-14380	-4952	7301	-2852
	(3193)	(3651)	(4814)	(6296)	(7718)	(8604)	(9698)	(11421)	(13247)	(15309)	(16812)
N	4556	4546	4539	4529	4321	3941	3542	3070	2599	2116	1644
	Panel E: Earnings results for representative sample of other mothers with singletons and twins (intensive fertility margin)										
Number of children	-2515	-22624***	-19254***	13963**	12113	264	572	2056	-1965	-4918	-14404
	(2332)	(2644)	(3757)	(5555)	(7391)	(8456)	(9649)	(10850)	(12338)	(14420)	(17633)
N	105610	105291	104951	96948	89200	81622	74306	66819	59387	52064	44606

Notes: Panel A replicates the main IV findings of Table 3. Panel B shows the effect of having a child on earnings for childless women entering their fourth IVF treatment. Fertility is instrumented with indicator of succes at fourth treatment. Panel C show the effect for women with previous children entering their first IVF treatment. Family size is instrumented with indicator of succes at first treatment. Panel D shows the effect of family size on earnings for women who had a succesful first IVF treatment. Family size is instrumented with a an indicator of having a singleton versus twin birth. Panel E shows the corresponding results for a sample of representative, non IVF-treated, women. The separate columns show estimates from separate regressions. Time period t=0 refers to the year of the (potential) child birth. All regressions control for age at IVF treatment, year of IVF treatment, and pre-treatment education and earnings. Standard error are in parentheses* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendices for Online Publication

Appendix A: Summary statistics for different samples

In this paper we use five different samples in estimation: (i) sample of IVF treated women without children (baseline sample); (ii) sample of IVF treated women with children; (iii) sample of IVF treated women without children entering fourth treatment; (iv) sample of successfully treated IVF women (with at least two implants for the twin strategy); and (v) sample or representative women. Table A provides sample means and standard deviations for some of the variables we use, together with sample sizes.

Appendix B: Balancing results

As our main findings, we report estimates of labor supply responses by year on unbalanced samples that vary by year. The unbalanced panel structure, which arises because information on long run labor market outcomes is not available for women who recently entered IVF treatment, may cause sample selection bias in our labor response estimates. In Table B we reestimate our main fertility and labor supply models on balanced samples that are fixed on groups of four consecutive treatment years. Our estimates are remarkably similar across the different but balanced samples, which indicate that the main estimates in Tables 3, 4 and 5 are not sensitive to our use of unbalanced samples.

Appendix C: OLS results

We also ran standard OLS regressions of annual labor earnings on having children, years of education and age using the two samples of women. One sample contains all IVF treated women. The other sample contains a 30 percent random sample of all other women who had their first child around the same time IVF treated women had their first IVF attempt. Table C reports these results. 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. Because long run labor market outcomes is not available for women who recently gave birth, samples sizes are falling over the years. While these least squares estimates already indicate that having children reduces annual labor earnings, the instrumental variable estimates using success at first IVF attempt as instrument are much larger and indicate that labor market consequences of first born children are underestimated when endogeneity of having children is ignored.

Appendix D: IVF related health risks

One concern in interpreting our results is that unsuccessfully treated IVF women are less healthy and, because of that, work fewer hours and receive lower salaries. We consider two possible scenarios. One is that IVF treatments are selective; that is, treatment success is less likely among women with unfavorable (and unobservable) health endowments. Another is that IVF treatments carry heterogeneous (mental) health risks; that is, treatment induces larger (mental) health risks among unsuccessfully treated women. Under both scenarios, we would underestimate the true fertilty effects because unsuccessfully treated IVF women face health risks that hinder their labor market performance.

To address these health concerns in more detail, we present results of three additional regressions that are (somehow) informative about unhealthy women, treatment failure, and poor labor market performance. Table D reports these estimates. In panel A we run our main instrumental variable regression, this time with additional controls for the number of eggs collected and transferred, diagnoses, causes of infertility, type of IVF treatment, and clinic indicators. Some of these variables will reflect potential health risks at the time of treatment. If the clinic was able to only collect a small number of eggs from a woman, this may reveal something about a woman's egg-producing potential and therefore reproductive health. Our estimates do not change much when we use this extensive set of controls. An alternative way of addressing the concern that failed IVF treatments are correlated with poor health, and thereby also to poor labor market performance, is to exploit the information on the type of infertility problem. In 40 percent of the couples, we observe that the source of the fertility problem lies in the male partner, for instance resulting from low sperm count or low sperm quality. In panel B we run our main instrumental variable regression on a restricted sample of women who do not experience any observed infertility problem themselves, but where the infertility problem is on the partner's side. Using this subsample, we are much less worried that women with poor (reproductive) characteristics drive the results. Again, we find that moving to a restricted sample of a priori healthy women has little consequence for our estimates; if anything, the estimates are somewhat larger in magnitude but less precise, due to the smaller sample size. In panel C we consider IVF related (mental) health risks and test whether women after a failed treatment face higher health risks (and get depressed). In there, we report estimates of the effect of success at first IVF treatment on registered intake of antidrepressants 0-10 years after the (potential) birth of the child. Apart from a small difference in the use of antidepressants in the year of treatment, our estimates indicate that women who seek IVF treatment face comparable mental health risks, regardless of treatment success. We all take this as evidence that health risks (related to failed IVF attempts) are not our biggest concern.

Table A: Descriptive statistics.

	(1)	(2)	(3)	(4)	(5)
Age at first treatment (or birth)	32.18	34.97	33.92	31.52	28.28
	(4.319)	(4.078)	(4.116)	(3.914)	(4.301)
Year at first treatment (or birth)	2000.2	2000.5	2000.4	2000.0	2001.4
	(3.107)	(3.107)	(2.997)	(2.989)	(4.066)
Pre-treatment labour income in	230.7	214.7	248.9	229.7	183.71
1000s DK (years 1-3)	(129.6)	(137.9)	(133.9)	(121.7)	(122.0)
Pre-treatment years of schooling	12.83	12.50	12.83	12.81	12.55
	(2.340)	(2.476)	(2.375)	(2.299)	(2.326)
Pre-treatment incidence of	0.169	0.191	0.261	0.172	0.143
sickness benefits	(0.375)	(0.393)	(0.439)	(0.377)	(0.350)
Married year before treatment	0.522	0.652	0.618	0.528	0.306
	(0.500)	(0.476)	(0.486)	(0.499)	(0.461)
Pre-treatment labour market	0.912	0.873	0.913	0.922	0.900
participation	(0.283)	(0.333)	(0.282)	(0.269)	(0.301)
Pre-treatment full time employment	0.865	0.834	0.884	0.880	0.773
	(0.342)	(0.372)	(0.320)	(0.325)	(0.419)
Observations	18586	4598	4299	4556	105610

Note: 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 IVFtreatment. Column (3): childless women entering their fourth IVF treatment. Column (4): women with a successful first IVF treatment attempt who had at least two embryos inserted. Column (5): representative sample of women with at least one child. Standard deviations within parentheses.

Table B: Fertility effects on female labor earnings: Results from instrumental variable regressions using balanced panels.

Independent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	t=0	t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10
	Panel A: 1995-1998										
Any children (0/1)	-46728***	-99030***	-27969***	-35164***	-36252***	-24124*	-35796**	-22636	-39973**	-36158**	-38661**
	(3100)	(5200)	(7239)	(9553)	(11948)	(13512)	(14515)	(15678)	(16890)	(18302)	(18808)
N	5052	5052	5052	5052	5052	5052	5052	5052	5052	5052	5052
	Panel B: 1999-2002										
Any children (0/1)	-48676***	-94436***	-11828*	-49887***	-43080***	-29187**	-42151***	-32408*			
	(2991)	(4946)	(7081)	(9579)	(11510)	(13430)	(15576)	(17627)			
N	6793	6793	6793	6793	6793	6793	6793	6793			
	Panel C: 2003-2006										
Any children (0/1)	-58980***	-102543***	-5486	-47718***							
	(3388)	(6091)	(9152)	(11670)							
N	6215	6215	6215	6215							

Notes: The table shows instrumental variables estimates of the effect of (potential) child birth on labor earnings for a set of balanced panels. The results are for a set of four-year windows, starting with those undergoing an IVF treatment in 1995-1998 and ending with those in 2003-2006. Note that the number of follow-up periods vary since earnings are observed in 2011 at latest. All models control for age at IVF treatment, year of treatment, pre-treatment education and earnings. Time period t=0 refers to the year of the (potential) child birth. Standard errors are in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C: External validity: Results from naive least squares regressions of annual labor earnings on having children.

Independent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	t=0	t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10
	Panel A:OLS results for childless women entering first IVF treatment										
Any children (0/1)	-40881***	-67369***	-24854***	-22341***	-15500***	-10147***	-9860***	-7602***	-7451***	-4507**	-2733
	(1312)	(1342)	(1378)	(1440)	(1459)	(1501)	(1625)	(1790)	(1957)	(2191)	(2452)
N	184959	184681	184337	183924	183440	182860	170546	156452	140731	123543	104947
	Panel B: OLS results for representative women										
Any children (0/1)	-43541***	-58770***	-22856***	-30563***	-24580***	-14949***	-12626***	-12816***	-13411***	-14400***	-13198***
	(506)	(514)	(534)	(556)	(564)	(576)	(587)	(631)	(678)	(736)	(809)
N	528963	528765	528377	527809	527013	525956	524459	489486	448734	402946	352501

Notes: Panel A shows the OLS relationship between having a first child and labor earnings for the sample of IVF treated women. Panel B shows the corresponding OLS relationship 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 in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D: IVF related health risks: Additional robustness results.

Independent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	t=0	t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10
Panel A: Earnings results with medical controls added to regression											
Any children (0/1)	-50597***	-96677***	-12191***	-41178***	-29028***	-18199**	-32765***	-21254*	-26529**	-13220	-40818**
	(1832)	(3173)	(4573)	(5986)	(7462)	(8963)	(10424)	(12140)	(13524)	(15366)	(17365)
N	18538	18494	18435	18381	17404	15599	13779	11983	10173	8342	6620
Panel B: Earnings results for a sample of healthy mothers											
Any children (0/1)	-50490***	-98456***	-5701	-41724***	-29206**	-31130*	-51469***	-27273	-39872	-24851	-45670
	(2814)	(5079)	(7720)	(10265)	(13113)	(16178)	(19502)	(22734)	(25943)	(27905)	(31891)
N	7378	7352	7330	7304	6872	6057	5228	4438	3691	2943	2250
Panel C: Consumption of anti-depressants											
IVF success (0/1)	-0.02***	-0.01	-0.01	-0.00	0.00	0.01	-0.00	0.00	-0.00	-0.00	-0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
N	18538	18494	18435	18381	17404	15599	13779	11983	10173	8342	6620

Notes: Panel A reruns our main specification (panel A in Table 3) but controls for the following medical factors: number of eggs collected, number of embryos implanted, diagnoses, causes of infertility, type of IVF treatment, and clinic indicators. Panel B shows the effect of a having a child on labor earnings for a sample of healthy women. The latter is defined as women who are *not* diagnosed with any fertility problem and where the fertility problem is on the partner's side. Panel C shows the reduced form effect of success at first IVF treatment on the probability of using anti-depressants. Time period t=0 refers to the year of the (potential) child birth. All regressions control for age at IVF treatment, year of treatment, and pre-treatment education and earnings.Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.