I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16959

# Is There Really a Child Penalty in the Long Run? New Evidence from IVF Treatments

Petter Lundborg
Erik Plug
Astrid Würtz Rasmussen

APRIL 2024

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16959

# Is There Really a Child Penalty in the Long Run? New Evidence from IVF Treatments

**Petter Lundborg**
*Lund University and IZA*

**Erik Plug**
*University of Amsterdam, Tinbergen Institute, IZA and UCLS*

**Astrid Würtz Rasmussen**
*Aarhus University*

APRIL 2024



ISSN: 2365-9793

**IZA – Institute of Labor Economics**

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

# ABSTRACT

## Is There Really a Child Penalty in the Long Run? New Evidence from IVF Treatments*

Newly matched data on in vitro fertilization (IVF) treatments are used to estimate the long-run consequences of children on the labor market earnings of women and men (often referred to as child penalties). We measure long-run child penalties in IVF-treated families by comparing the earnings of successfully and unsuccessfully first-time treated women and men up to 25 years after the first IVF treatment. In the short run, we find a large penalty immediately after the birth of the first child. In the long run, however, we find that the child penalty fades out, disappears completely after 10 years, and even turns into a child premium after 15 years, offsetting the initial setbacks experienced when children are young. Our findings therefore challenge the widely held view that children are the primary drivers behind the long-run gender gap in earnings.

**JEL Classification:** J13, J16, J31

**Keywords:** child penalty, gender earnings gap, fertility

**Corresponding author:**
Erik Plug
Amsterdam School of Economics
University of Amsterdam
P.O. Box 15867
1001 NJ, Amsterdam
The Netherlands

E-mail: e.j.s.plug@uva.n

* We thank Gordon Dahl, Erik Lindqvist, Oskar Nordström Skans, Johanna Rickne, Núria Rodríguez-Planas, Jacob Søgaard, and seminar/conference participants in Amsterdam, Belgrade, Copenhagen, Essen, Lenzerheide, London, Lund, and Prague 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). DNBH is not responsible for the analysis or interpretation of the data presented. Financial support from Riksbankens Jubileumsfond is gratefully acknowledged.

Explaining why women earn persistently less than men continues to challenge social scientists. One explanation, shared by many economists, points to children as primary cause (Goldin 2014, Bertrand 2020, Cortés and Pan 2023). It builds on the common observation that women, not men, take care of children. In the short run, when children are young, women earn less than otherwise similar men because they, as mothers, prioritize child care and work fewer hours, quit their jobs, or change to lower paid jobs closer to home. In the long run, when children get older and need less care, the same women continue to earn less because they, as primary caretakers, lost relevant labor market experience, missed out on earnings opportunities, or ended up in more child-friendly but lower paid jobs with flatter earnings profiles.[1]

Testing whether children are the primary cause for persistent gender earnings gaps has proved difficult, however. More specifically, almost all previous research on child penalties - these are the impacts of parenthood on the earnings of women and men - are more informative about short-run than long-run penalties because of methodological limitations and data constraints.

There are broadly two approaches to identify child penalties. The commonly used event studies identify short- and long-run child penalties by comparing the earnings of mothers of the same age who timed their first births either close or far apart (Kleven et al. 2019 2023, Cortés and Pan 2023). Because the identifying assumption of random timing of births is less plausible for mothers who had their first births spaced far apart, event study designs are less suited for the estimation of long-run child penalties (Goldin et al. 2022, Bensness et al. 2023). The less commonly used instrumental variable (IV) studies based on in vitro fertilization (IVF) treatments identify short- and long-run penalties by effectively comparing the earnings of successfully and unsuccessfully first-time treated women either shortly or long after their first treatment (Lundborg et al. 2017, Bensness et al. 2023, Räsänen 2023, Ilciukas 2024). Because the administrative registers with IVF information became more recently available, the estimation of child penalties is mostly limited to women who were first treated about 5-10 years ago. Given that labor market careers last approximately 30-45 years after the birth of the first child, it is fair to say that we know little to nothing about long-run child penalties and their impact

[1]There is a large and related literature on family-friendly policies and gender earnings gaps. These policies take the career costs of children as a mere fact and aim to lower the child care burden through, among others, compensated parental leave arrangements and subsidized child care programs (e.g., Lalive and Zweimüller 2009, Havnes and Mogstad 2011, Dahl et al. 2016).

on long-run gender earnings gaps.

In this paper, we provide new evidence on the long-run child penalty by analyzing data from Danish women who underwent IVF treatments. We take the approach proposed by Lundborg et al. (2017) and compare the labor market outcomes of successfully and unsuccessfully treated women (and their partners) for up to 25 years following their first IVF treatment. Since observed working histories of successfully and unsuccessfully treated women (and men) are virtually identical before they enter their first IVF treatment, this comparison informs us how important children are in expaining gender earnings gaps over a substantial share of the women's and men's career.

With administrative data from Denmark, our findings challenge the conventional wisdom that children are the key drivers behind long-term gender earnings gaps. We find a large child penalty shortly after the birth of the first child. But we also find that this penalty fades out over time, with mothers eventually catching up to their childless counterparts. The child penalty has disappeared after 10 years, and even turns into a child premium after 15 years, offsetting the initial setbacks experienced during the early stages of motherhood. Our findings thus suggest that children contribute little to nothing to the persistent gender gap in earnings.

We continue our analysis by testing the validity of these absent penalties. We first consider several threats that could mask child penalties in our IV-IVF strategy (internal validity). In particular, we examine two possible violations of the exclusion condition. First, women, after a first failed treatment, may have children in later treatments. If child penalties are larger in the short run, we would underestimate child penalties in the longer run. We show that the impacts of delayed children are irrelevant in the long run. Second, women, after a first failed treatment, may be more at risk to stay childless and develop mental health problems. If these problems adversely affect their earnings, we would again underestimate the child penalty. We argue that, in the long run, most women have had enough time to settle their mental affairs. In addition, we show that mental health problems (measured by antidepressant medication) are empirically irrelevant (Lundborg et al 2017).

We next consider whether IVF-treated women are simply different women facing smaller career costs of children (external validity). Given that these women are, on average, richer, more educated, and older when they have their first child, they may be better in combining care and work. We test for this in two ways. First, we examine whether there are larger long-run child penalties for IVF-treated women who resemble the average representative mother (being less affluent, less educated, and younger). We find no evidence suggesting so. Second, we run an event study design on our sample of IVF-treated women. If IVF-treated women face smaller child penalties than other women, we should see this reflected in event study estimates. We do not; the event study estimates we obtain are comparable to those obtained using representative samples of Danish women.

And lastly, we contrast our IV-IVF estimates to those obtained using the event study design. Unlike the absent penalties in our IV-IVF setup, we find large and long-lasting child penalties in an event study design run on the sample of IVF-treated mothers. We consider several internal validity threats that could amplify penalties in event study designs. In particular, we study two possible violations of the parallel trend assumption. First, women who are younger at their first birth may experience flatter pre-birth earnings trends. We find no evidence of this. Second, women may time their fertility to periods when earnings profiles flatten, as argued by Bensnes et al. (2023). We find some suggestive evidence that childless women indeed face much flatter earnings profiles after their first treatment failed. Such flattening of earnings profiles is not captured by the event study design, and may lead to child penalties being too large.

In the end, we feel confident enough to conclude that children are not responsible for the long-run gender gaps in earnings. The mechanisms behind the persistent gender earnings gap should therefore be sought elsewhere.

# I IVF Context and Data

In this section we outline some facts about IVF treatments in the Danish context, shortly discuss the IVF register, and provide details about the IVF sample we use in our main analysis.

## A IVF treatment

IVF treatment is the leading medical intervention to help infertile couples become pregnant and conceive children. The majority of treatments is given to childless couples. With a practitioner's referral, childless couples are entitled to 3 free treatments. Without a practitioner's referral, childless couples must pay for their treatment. A typical IVF treatment package of 3 treatments goes for about 60.000 DKK. IVF treatments, which on their own are quite invasive, can also be costly in physical and mental terms.[2] Of all Danish children born between 1995 and 2005, 2 to 3 percent are IVF children.

## B IVF data

The Danish IVF register holds information on IVF treatments in all fertility clinics and hospitals. The register had full coverage between 1994 and 2005. It includes information on the reason for infertility, number of eggs retrieved from the womb, number of fertilized eggs transferred back, treatment outcome (birth, abortion, stillbirth, or failure), date of treatment, and where applicable date of birth. The IVF register is matched to other registers to get longitudinal information from 1991 to 2020 on education, age, gender, marital status, number of children and various labor market variables (including annual earnings). During our study period, the IVF register contains information on 31,666 women receiving altogether 96,807 IVF treatments.

## C IVF Sample

We conduct much of our analysis on the sample of childless couples in their first IVF treatment who successfully reached the final stage of embryos implants. This IVF sample consists of 18,547 treated women.[3] Online Table 1 contains descriptive statistics for some pre-treatment characteristics and post-treatment outcomes, by treatment success at first IVF treatment. We find that, on most dimensions, successfully treated women in the IVF sample are remarkably similar to unsuccessfully treated women. Importantly, the education levels and earnings of treated women and their partners are similar before they seek IVF treatment.[4]

[2]Each IVF treatment involves four stages: fertility medication used to stimulate egg development; egg collection; in vitro fertilization, where eggs and sperm are mixed in a laboratory setting; and the in-utero transfer of one or more selected embryos. In our main analysis, we focus on childless couples in their first treatment that reached the final stage and had embryo implants.

[3]We arrive at a sample of 18,547 treated women through the following selection rules. Of the initial 32,073 women treated between 1994 and 2005, we remove 2,908 in 1994 (to secure first treatments), 4,660 with children (to measure the onset of motherhood), 5,255 who had no eggs inserted because of failed egg production or fertilization (to select women who are as similar as can be), 260 with missing key controls (to test for balance), and 424 we could not match to the earnings register (to estimate child penalties).

[4]One dimension in which treated women significanty differ is age, where successfully treated women are, on average, older. This is in line with medical evidence suggesting that age is the single most important factor determining success in IVF treatments (Rosenwaks

Online Table A1 also contains descriptive statistics of a 30 percent representative sample of mothers who gave birth during the same time period 1995-2005. When we compare the treated women to those untreated women in the representative sample, we find that the IVF treated women are older, better educated, and more highly paid.

# II IV-IVF stategy

In our empirical analyses we mostly focus on reduced-form regressions on the effect of success at first IVF treatment on the labor market earnings of women. The reduced form specification can be written as:

$$(1) \quad Y_{it} = \beta_{1t} X_i + \beta_{2t} Z_i + u_{it},$$

where $Y$ represents the earnings of woman $i$ treated $t$ years ago, $X$ represents a set of exogenous control variables including the woman's age at first treatment, year of treatment, and years of education, $Z$ is the instrumental variable (1 if the first IVF treatment with embryo implants has lead to a child, and 0 otherwise), and $u$ represents the econometric error and contains unobservable factors which, conditionally on $X$, are assumed unrelated to $Z$. The parameter of interest is the reduced-form parameter $\beta_{2t}$, which captures the causal effect of IVF treatment success on earnings $t$ years after entering the first IVF treatment.

In supplementary analyses, we also run the following first- and second-stage regressions:

$$(2) \quad F_{it} = \alpha_{1t} X_i + \alpha_{2t} Z_i + \epsilon_{it},$$

$$(3) \quad Y_{it} = \gamma_{1t} X_i + \gamma_{2t} \widehat{F}_{it} + \nu_{it},$$

where $F_{it}$ is a fertility indicator (1 if woman $i$ has children $t$ years after treatment, and 0 otherwise), and $\epsilon$ and $\nu$ are the econometric errors, which contain unobservable factors that can be related to fertility and earnings. As before, the same unobservable factors are assumed unrelated to $Z$ (conditionally on $X$). In this setup, $Z$ is the instrument for $F_{it}$, and separate IV regressions

et al. 1995, Templeton et al. 1996, van Loendersloot et al. 2014). While not observed in our sample, another dimension in which treated women sometimes differ is education (Groes et al. 2017). Our analyses accounts for this by including age and education controls. It is important to note, though, that the penalties we estimate do not depend on whether we run regressions with or without controls (see Online Appendix Figures A5 and A6).

are run for each year after the first IVF treatment. The second-stage parameter $\gamma_{2t}$ (attached to the first-stage predicted fertility indicator) represents the child penalty and captures the causal effect of having children on female earnings $t$ years after entering the first IVF treatment. The plausibility of the identifying assumptions (related to instrument relevance, independence, and exclusion) will be discussed in the sections below.

# III Results

We first provide descriptive earnings patterns for successfully and unsuccessfully treated women and men, before and after their first IVF treatment. Figure 1 displays earnings from 6 years prior to first IVF treatment up to 25 years after. The solid lines show female earnings (the black line represents the earnings of successfully treated women). The dotted lines show male earnings (the long dots represent the earnings of men in successfully treated couples). The sample is balanced until 15 years after the IVF treatment. The sample becomes unbalanced thereafter.

The graph yields three insights. First, successfully and unsuccessfully treated women follow remarkably similar earnings trajectories in the years leading up to the first IVF treatment. Not only do their trends closely mirror each other, they also display similar earnings in levels. Online Appendix Table 1, which contains sample means and standard deviations for several pre-treatment characteristics and post-treatment outcomes, confirms that both groups have almost identical pre-treatment earnings. The close similarity between successfully and unsuccessfully treated women, up to their first treatment, suggests that treatment success itself does not depend on the women's pre-treatment labor market skills and earnings. Shortly after their first treatment, however, the earnings of successfully treated women drop substantially when compared to their unsuccessfully treated counterparts. This is a first sign of a substantial short-run child penalty. We see that within two years the earnings of successfully treated women already start to recover and converge towards the earnings of their unsuccessfully treated counterparts. By year 10, successfully treated women earn as much as unsuccessfully treated women. And by year 15, successfully treated women earn slighly more. This earnings advantage persists throughout the remainder of the study period.

Second, the men in successfully treated couples display similar earnings trends and levels as the men in unsuccessfully treated couples in the years leading up to the first IVF treatment. As opposed to the fall in female earnings, however, we see no dip in male earnings. Instead, both groups of men continue to closely track each other's earnings in the years following the first IVF treatment as if nothing has happened. Towards the end of the study period, the male earnings for both groups fall, which we attribute to the rising share of retired men.

Third, gender differences in earnings (in IVF couples) are already present before the first IVF treatment attempt. In the years leading up to the initial IVF procedure, women earn approximately 20 percent less than their male counterparts. This is in line with findings from numerous other studies, where gender differences in fields of study, occupational choices, and labor supply (measured in hours) can explain much of the gender differences in earnings that manifest before parenthood (Cortés and Pan, 2023).

## A Fertility effects

For our empirical analyses to make sense, it is important that IVF success predicts long-run fertility. If all IVF-treated women end up having children, we would not be able to estimate the long-run child penalty. Figure 2 (panel A) shows that treatment success at the first IVF attempt strongly affects fertility in the short, medium, and long run. In the short run, treatment success exerts a strong and immediate impact on the likelihood of motherhood. Since some women who fail at their first treatment continue with additional treatments in the same year, where some women experience successful IVF treatments, the coefficient is less than one. In subsequent years, the impact of success at first IVF treatment on motherhood decreases in magnitude, as many women undergo successful IVF treatments at later attempts. Nonetheless, the long-run effect of IVF treatment success on motherhood remains positive and significant, implying that childless women whose first IVF treatment did not lead to pregnancy and childbirth are also more likely to remain childless in the long run.

## B Long-run child penalties

We next turn to the main questions of the paper. First, is there a child penalty in long run? Second, is the child penalty one of the primary drivers behind the persistent gender gap in earnings?

For the first question, we run regressions using female and male annual earnings as dependent variables and IVF success at first birth as primary independent variable, controlling for female and male years of education, age at first treatment, and year of treatment, respectively. In these regressions, the year before the (potential) birth of a child serves as the omitted category. Figure 2 (panels B and C) plot the corresponding reduced-form regression estimates of the effect of IVF success on female and male earnings for each year, spanning from 6 years prior to the first IVF treatment to 25 years thereafter. We see that pre-treatment earnings trends are strikingly similar across successfully and unsuccessfully treated women and men, confirming that IVF success does not depend on women's and men's earnings before their first IVF treatment. We also see that, after the first IVF treatment attempt, there is an impact of IVF success on female earnings, not on male earnings, confirming the common notion that women, not men, bear the costs of child care. The effects are most apparant in the years shortly after the IVF treatment, when successfully treated women experience a sharp decline in earnings. Two years later, however, succesfully treated women already begin to recover; 10 years later, succesfully treated women are fully recovered and earn as much as their unsuccessfully treated counterparts; and from that point onwards, the successfully treated women earn as much as (if not more than) than unsuccessfully treated women.[5]

Since the working life of the women in our sample spans another 30-35 years after their first treatment, we have also calculated the child penalty from a life-cycle perspective. The successfully treated women are, on average, 32 years old at first treatment and thus have 33 working years left before they retire at 65. If we ignore discounting, assume a working life of 40 years (running from 25 to 65), and fix the unknown penalty in the 8 years before retirement at the penalty we estimate 21-25 years after child birth (derived from the first-stage and reduced-form estimates in Online Appendix Table 2, columns 1 and 2), we find that the birth of the first child does not lower lifetime female earnings. If anything, we calculate that motherhood leads to a small rise in lifetime female earnings cycle of 2-3 percent. Based on these findings, we conclude that there is no evidence of a sizable long-run child penalty.

For the second question, we run the same regression but switch the depen-

[5]These near zero long-run child penalties are also reflected in the other labor market outcomes of IVF-treated women. Online Appendix Figure A1 shows no systematic long-run differences between the labor force participation, hours worked, and hourly wages of successfully and unsucessfully treated women.

dent variable to the within-couple gender gap in annual earnings and control for years of education and age at first treatment of both women and men in the IVF-treated couple. Figure 2 (panel D) plots the reduced-form estimates. We see that children contribute little to nothing to the gender earnings gap in the long-run. While within-couple earnings gap sharply grows shortly after the first treatment, the same gender gap quickly reverts to its pre-treatment level within 8 years following the first IVF treatment. We find similar patterns when we consider the share of household income earned by women (see Online Appendix Table A2). Based on these findings, we conclude that children have little to do with long-run gender gaps in earnings.

## C Robustness

The estimates plotted in Figure 2 are taken from unbalanced samples and regression models that control for years of education, year of treatment and age at first treatment. We obtain similar estimates for balanced samples, for samples without twin births, for samples restricted to healthy women with less fertile partners, and for regression specifications without and with a full set of control variables (see Online Appendix Figures A2, A3, A4, A5 and A6).

# IV Internal validity threats

Our findings, so far, suggest that there is no long-run child penalty. To make sure that these absent penalties are real, we must ask ourselves whether there are any internal validity threats that could possibly mask sizable child penalties using our IV-IVF strategy. In particular, we consider two likely violations of the exclusion condition (which dictates that success at first treatment impacts earnings exclusively through its impact on fertility). For the analysis to follow, we focus on the earnings of women as our main labor market outcome.

*Delayed fertility.* Women, after a first failed treatment, may have children in later treatments. If penalties are larger in the shorter run, we would underestimate the penalty in the longer run. To illustrate how children from later treatments bias the penalty, we introduce a simplified three-period fertility model with two groups of women who experience different short- and long-run child penalties. The first group consists of compliers: these women remain childless after a first failed IVF attempt. They earn $Y_c$ in the absence of children. After a successful first treatment, they experience an earnings loss

$C_c$ in period 1 when children are young, and $c_c$ in periods 2 and 3 when children are older. The complier share is set to $\beta_c$. The second group consists of always takers. These women will always end up having children regardless of the treatment outcome at the first IVF attempt. They earn $Y_a$ in the absence of children. After a first successful treatment (in either period 1 or 2), they experience an earnings loss $C_a$ when children are young, and $c_a$ when children are older. The always-taker share is set to $\beta_a$. We note that, by construction of the instrument, our IVF population only consists of compliers and always takers ($\beta_c + \beta_a = 1$). Within this simplified fertility model (without any other covariates), we can derive the corresponding first-stage and reduced-form estimates for all three periods under the assumptions that treatment success at first treatment is effectively random and only impacts earnings through its impact on children.

Table 1 contains the theoretical first-stage and reduced-form effect estimates. The first-stage estimate in period 1 equals 1. It represents all successfully first-time treated women, which consists of compliers and always takers that had a child in period 1 ($\beta_c + \beta_a$). Given that children are costly, the reduced-form effect estimate in period 1 represents the earnings losses among the compliers and successfully treated always takers weighted by their population shares ($-\beta_c C_c - \beta_a C_a$). The second-stage child penalty is the reduced-form estimate divided by the first-stage estimate. In a related vein, the first-stage estimate in period 2 only consists of compliers ($\beta_c$), as all the always takers have had their children by now. The reduced-form estimate in period 2, however, captures the causal effect of interest ($-\beta_c c_c$) and some nuisance parameter ($\beta_a[C_a - c_a]$), which represents the delayed fertility impact among the always takers with a failed treatment in period 1. Because child penalties are assumed larger when children are young ($C_a > c_a$), delayed fertility reduces the average earnings of always takers and attenuates the estimated child penalty in period 2. And lastly, in period 3, the first-stage and reduced-form effect estimates no longer depend on always takers (because their earnings losses are assumed similar in periods 2 and 3) and exclusively capture the complier share ($\beta_c$) and the long-run impact of treatment success at the first IVF attempt on their earnings ($-\beta_c c_c$). Together, these estimates uniquely identify the long-run impact of children on the earnings of women who were treated successfully at their first IVF attempt.

In sum, the child penalties we estimate represent a mixture of fertility and delayed fertility effects on labor earnings. In our simple model, delayed fertility will be a concern in period 2 (medium run) when it biases the child penalty towards zero. In more general terms, the long-run impact of delayed fertility will cause little concern when children age and child penalties get smaller, and converge to some fixed amount.

*Mental health problems.* Women, after a first failed treatment, may be disappointed. If disappointment and eventual childlessness, after a sequence of failed treatments, triggers some emotional response that lowers earnings, the child penalty not only captures the child-driven earnings loss of successfully treated women but also the emotion-driven earnings loss of unsuccessfully treated women. Since the penalty is effectively measured by the earnings difference between successfully and unsuccessfully treated women, we would again underestimate the child penalty.

We argue that such a bias will be small in the long run. First, the medical evidence on the (long-run) impact of IVF treatments on female depression is at best weak. With self-assessed depression scores, most survey studies find that treated women adjust well to unsuccessful IVF treatments (Verhaak et al. 2007). With medical depression indicators (antidepressant medication, psychiatry visits, or hospitalizations for mental disorders), registered-based studies find that depression rates are generally low among IVF treated women, yet slightly higher among never successfully treated women (Agerbo et al. 2013, Baldur-Feskov et al. 2013, Pedro et al. 2019, Yli-Kuha et al. 2010). Second, we estimate in earlier work the impact on IVF treatment success on antidepressant medication (Lundborg et al. 2017). Based on the same IVF sample we use in this study, we follow women up to 10 years after their first treatment, and find that, 2 years later, the remaining effects of IVF success on prescribed antidepressant medication are all small and statistically insignificant. And third, a simple calculation suggests that the bias introduced by mental health problems can only be small. To do this, we take mental health penalties from elsewhere and attribute earnings losses of 10, 20, and 40 percent to those with mild, moderate, and severe mental health problems (Bartel and Taubman 1986, Frank and Gertler 1991). If we assume, conservatively for us, that 20, 10, and 5 percent of never successfully treated women experience some mild, moderate, and severe form of depression, the long-run bias (calculated by the average long-run earnings loss of unsuccessfully treated compliers) can mask child penalties of at most 6 percent.

# V External validity threats

The next question we ask is whether the absent long-run child penalties in samples of IVF treated women are informative about the larger population of representative mothers. There are many reasons why IVF treated women may face lower career costs of children than other women: they earn more and find it easier to outsource child care; they are more educated and, as more informed decision makers, are better in combining care and work; they are older and, with more realized careers, find it easier to deal with children; they have fewer children and miss out on penalties of subsequent children; they express a stronger demand for children and, as such, put less value on their career; and so on.

To detect whether IVF treated women are indeed the type of women with lower career costs of children, we perform two tests. First, we apply our IV-IVF strategy and test whether there are larger long-run child penalties for IVF-treated women who resemble the average representative mother (being less affluent, less educated, and younger). And second, we run an event study design on our sample of IVF treated mothers and test whether they face smaller child penalties than the other, more representative, mothers.

Figure 3 plots all the different child penalty estimates for the different samples and strategies. When we look at the IV-IVF child penalties for IVF-treated women, which we take as our point of reference, we find as before that the short-run child penalties quickly disappear when children age. When we look at IV-IVF child penalties for those IVF-treated women similar to the mothers in the representative sample, we find the same dissappearing child penalties when children get older.[6] We next turn to the event study analysis run on our sample of IVF treated women. We depart here from our IV-IVF design and instead compare the earnings of similarily aged women who were all successfully treated and became mothers at difference points in time. Our event study analysis controls for year, time, and age effects, as in Cortés and Pan (2023). We restrict the analyses to a 10 year follow-up period, since it makes little sense to assume random timing of births when spaced more than 10 years apart. When we look at the event study estimates for

[6] The sample of IVF treated women are reweighted to match the sample of representative mothers. The weights are the inverse propensity scores taken from a probit regression of being a representative mother on the child's birth year (which we calculate for unsuccessfully treated women by adding nine months after the day of embryo transfers), years of schooling, pre-treatment earnings, and birth year. For common support reasons, we discard 121 observations with propensity scores below 0.01 and above 0.99.

the IVF-treated mothers, we find an immediate decline in earnings following childbirth. While there seems to be a slight recovery after two years, earnings remain consistently lower throughout the observation period. We find that 10 years after childbirth, the penalty for IVF-treated mothers amounts to approximately 20 percent. When we look at the event study estimates for all other mothers, we see large and persistent penalties, as in Kleven et al. (2019a). We also see that the long-run child penalties differ only a little from those obtained for IVF-treated mothers. Altogether, we find little evidence that, when it comes to the career costs of children, IVF-treated women are very different than most other women.

# VI Why do child penalties differ?

Figure 3 also highlights sharp differences in the long-run child penalities between the IV-IVF and event study designs; while the IV-IVF estimates show a fading out of penalties, the event study estimates show large and persistent penalties. Why is this? If we take the IV-IVF child penalties as real, there must be some systematic bias in an event study design that amplifies child penalties in the long run. In particular, we consider two likely violations of the parallel trend assumption (which dictates that mothers who already had their children would, in the absence of children, earn as much as other mothers similar in age and birth cohort who did not had their children yet). For ease of comparison, we focus on the penalty estimates obtained with IVF-treated mothers (to ensure that the differences between IV-IVF and event-study estimates are not driven by the differences in the population studied).

*Different pre-trends.* Mothers who are younger at their first birth may experience flatter pre-birth earnings trends. If we subsequently base their counterfactual earnings on the earnings of older mothers with steeper pre-child earnings profiles, we would overestimate the child penalty. Figure 4 (panel A) plots the earnings trajectories of women by age at motherhood. Here, we split the sample into 4 age groups; below 26, 26-30, 31-35, and 36+. The steepest pre-birth earnings profile is found among mothers aged 26-30, while the flattest one is found for women aged 36 and above. Since earnings profiles are not flatter among younger mothers, we can rule out that different pre-trends in earnings bias the child penalty estimates upwards.

*Different trend breaks.* Young and older mothers may time their fertility when their earnings profiles start to flatten, as suggested by Bensnes et al.

(2023). If we base the counterfactual earnings of young mothers on only the pre-child earnings of older mothers, and ignore that their earnings profile would flatten, we would again overestimate the child penalty. Figure 4 (panel B) indeed suggests that women time their births when earnings profiles flatten out. In there, we compare the earnings of women who succeed in their first IVF treatment to those of women who fail their first IVF treatment and do not have children yet (that is, women with a failed first treatment are only included up to the point they have their child). Similar to findings from Norway (Bensnes et al. 2023), we see that the latter group of women experience much flatter earnings profiles, after their first treatment, in the absence of children. If this flattening out of post-treatment earnings would also happen to successfully treated women in the counterfactual without children, event studies would overestimate child penalties in the long run.

It is important to note that, in an IV-IVF strategy, such flattening out is taken into account. When we derive the penalty by comparing the earnings of women who attempt to have children at the same point in time, there will be no bias as long as those who succeed do so for random reasons (implying that successfully and unsuccessfully treated women, in the absence of children, are equally exposed to the same flattened earnings profiles). We conclude that such flattening out seems a plausible reason why long-run estimates between the two identication designs diverge.[7]

# VII Concluding remarks

The child penalty has been singled out as one of the primary drivers behind the gender gap in earnings. In this paper, we challenge this notion by estimating the child penalty in the very long run. For this purpose, we rely on an instrumental variable strategy based on IVF-induced fertility variation among childless couples in Denmark to identify child penalties for up to 25 years after

[7] One possible concern is that the flattening out of post-treatment earnings, observed after the first failed treatment among never successfully treated women, occurs because childless women already earn less in anticipation of children. While the impact of anticipated children would be captured by the event study estimates but missed by IV-IVF estimates, we argue that anticipation effects on the long-run child penalties can only be small. In particular, the impact of anticipated children must be unrealistically large (and much larger that the impact of realized children) to explain both the large penalties in event study designs (representing a mixture of anticipated and realized effects of having children on earnings) and the zero penalties and/or premiums in IV-IVF strategies (representing the realized effect of having children on earnings).

the birth of the first child. With this IV-IVF strategy, we find that the first child impacts the earnings of women, not men. While the child penalties are sizable shortly after birth, the same penalty fades out, disappears completely after 10 years, and turns into a child premium after 15 years. When we evaluate the child penalty from a life-cycle perspective, we even find that the birth of the first child leads to a small rise in the lifetime earnings of women. On the whole, we find no evidence that children can explain why women earn persistently less than men.

Our findings are in sharp contrast with the sizable long-run child penalties found in event studies. To explain the differences, we argue that event study designs are less suited to estimate long-run child penalties particularly when women time their fertility when their earnings start to flatten out. To explain why sizable short-run child penalties quickly fade out, we argue that children primarily affect their mother's earnings during the years mothers take care of their children. As children grow older and demand less care, we see that the mother's earnings start to recover, with much of the immediate penalties made up 10 years after the birth of the first child.

# References


**Agerbo, Esben, Preben Bo Mortensen, and Trine Munk-Olsen**, "Childlessness, Parental Mortality and Psychiatric Illness: A Natural Experiment based on In Vitro Fertility Treatment and Adoption," *Journal of Epidemioligy and Community Health*, 2012.

**Baldur-Felskov, Birgitte et al.**, "Psychiatric Disorders in Women with Fertility Problems: Results from a Large Danish Register-Based Cohort Study," *Human reproduction*, 2013, *28* (3), 683–690.

**Bartel, Ann and Paul Taubman**, "Some Economic and Demographic Consequences of Mental Illness," *Journal of Labor Economics*, 1986, *4* (2), 243–256.

**Bensnes, Simon, Ingrid Huitfeldt, and Edwin Leuven**, "Reconciling Estimates of the Long-Term Earnings Effect of Fertility," IZA Discussion Papers 16174, Institute of Labor Economics (IZA) May 2023.

**Bertrand, Marianne**, "Gender in the Twenty-First Century," *American Economic Review: Papers and Proceedings*, 2020, *110*, 1–24.

**Cortés, Patricia and Jessica Pan**, "Children and the Remaining Gender Gaps in the Labor Market," *Journal of Economic Literature*, 2023, *61* (4), 1359–1409.

**Dahl, Gordon, Katrine Løken, Magne Mogstad, and Kari Vea Salvanes**, "What Is the Case for Paid Maternity Leave?," *The Review of Economics and Statistics*, 2016, *98* (4), 655–670.

**Frank, Richard and Paul Gertler**, "An Assessment of Measurement Error Bias for Estimating the Effect of Mental Distress on Income," *The Journal of Human Resources*, 1991, *26* (1), 154–164.

**Goldin, Claudia**, "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, 2014, *104* (4), 1091–1119.

**_ , Sari Pekkala Kerr, and Claudia Olivetti**, "When the Kids Grow Up: Women's Employment and Earnings across the Family Cycle," Technical Report, National Bureau of Economic Research 2022.

**Groes, Fane, Daniela Iorio, Man Yee Leung, and Raül Santaeulalia-Llopis**, "Educational Disparities in the Battle Against Infertility: Evidence from IVF Success," *EUI WORKING PAPERS ECO*, 2017, pp. 1–42.

**Havnes, Tarjei and Magne Mogstad**, "Money for Nothing? Universal Child Care and Maternal Employment," *Journal of Public Economics*, 2011, *95* (11-12), 1455–1465.

**Ilciukas, Julius**, "Bounding Child Penalties in Fertility Treatments," Unpublished Manuscript, University of Amsterdam 2024.

**Kleven, Henrik, Camille Landais, and Gabriel Leite-Mariante**, "The Child Penalty Atlas," Technical Report, National Bureau of Economic Research 2023.

**_ , _ , and Jakob Egholt Søgaard**, "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, October 2019, *11* (4), 181–209.

**_ , _ , Johanna Posch, Andreas Steinhauer, and Josef Zweimüller**, "Child Penalties across Countries: Evidence and Explanations," *AEA Papers and Proceedings*, 2019, *109*, 122–126.

**Lalive, Rafael and Josef Zweimüller**, "How Does Parental Leave Affect Fertility and Return to Work? Evidence from Two Natural Experiments," *The Quarterly Journal of Economics*, 2009, *124* (3), 1363–1402.

**Lundborg, Petter, Erik Plug, and Astrid Würtz Rasmussen**, "Can Women Have Children and a Career? IV Evidence from IVF Treatments," *American Economic Review*, June 2017, *107* (6), 1611–37.

**Pedro, Juliana et al.**, "Infertility-Related Stress and the Risk of Antidepressants Prescription in Women: A 10-year Register Study," *Human Reproduction*, 2019, *34* (8), 1505–1513.

**Räsänen, Tapio**, "How Does Motherhood Affect Women's Careers? Causal Estimates using Medically Induced Ovulation Treatments," Unpublished Manuscript, University of Jyväskylä 2023.

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

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

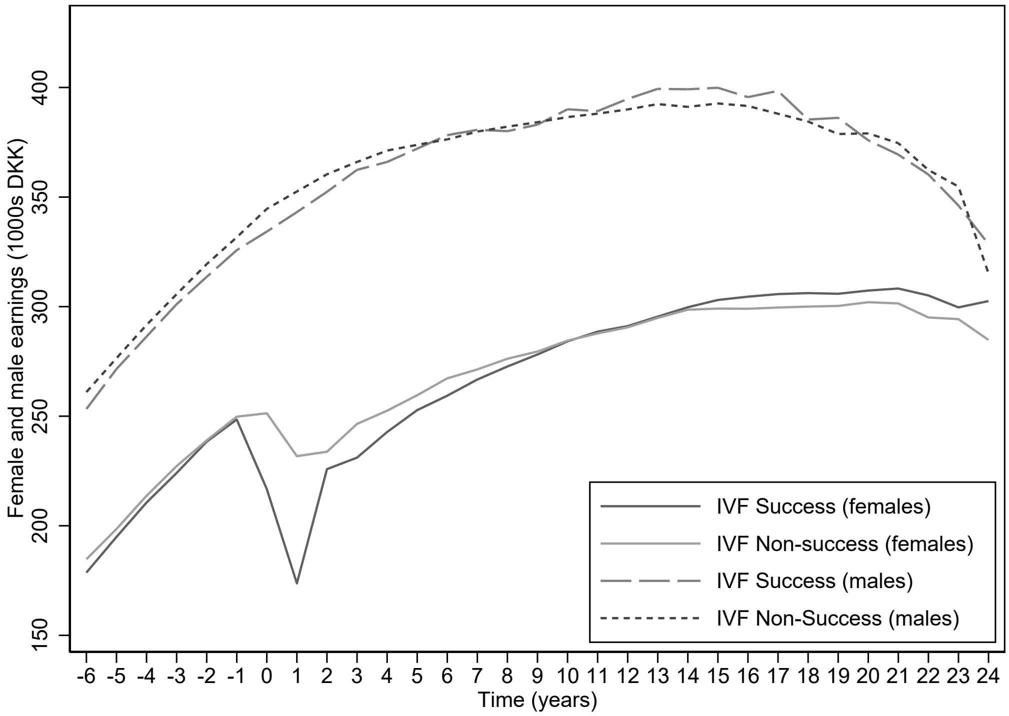
**van Loendersloot, Laura, Sjoerd Repping, Patrick Bossuyt, Fulco van der Veen, and Madelon van Wely**, "Prediction Models in *In Vitro* Fertilization; Where Are We? A Mini Review," *Journal of Advanced Research*, 2014, *5* (3), 295–301.

**Verhaak, Chris et al.**, "Women's Emotional Adjustment to IVF: A Systematic Review of 25 Years of Research," *Human Reproduction Update*, 2007, *13* (1), 27–36.

**Yli-Kuha, Anna-Niina et al.**, "Psychiatric Disorders Leading to Hospitalization Before and After Infertility Treatments," *Human Reproduction*, 2010, *25* (8), 2018–2023.

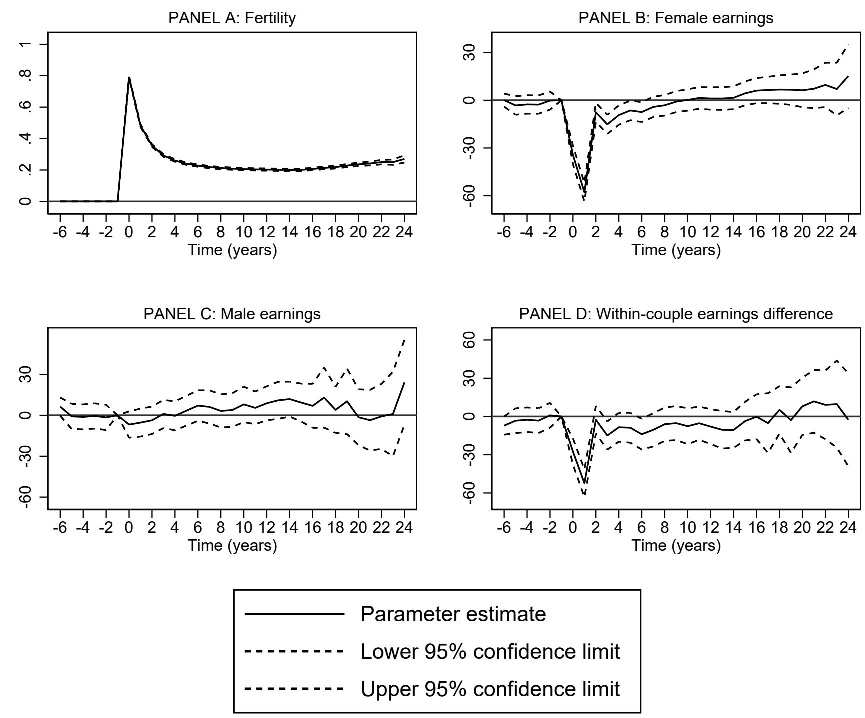
# Tables and Figures

Figure 1: Annual earnings before and after first IVF treatment



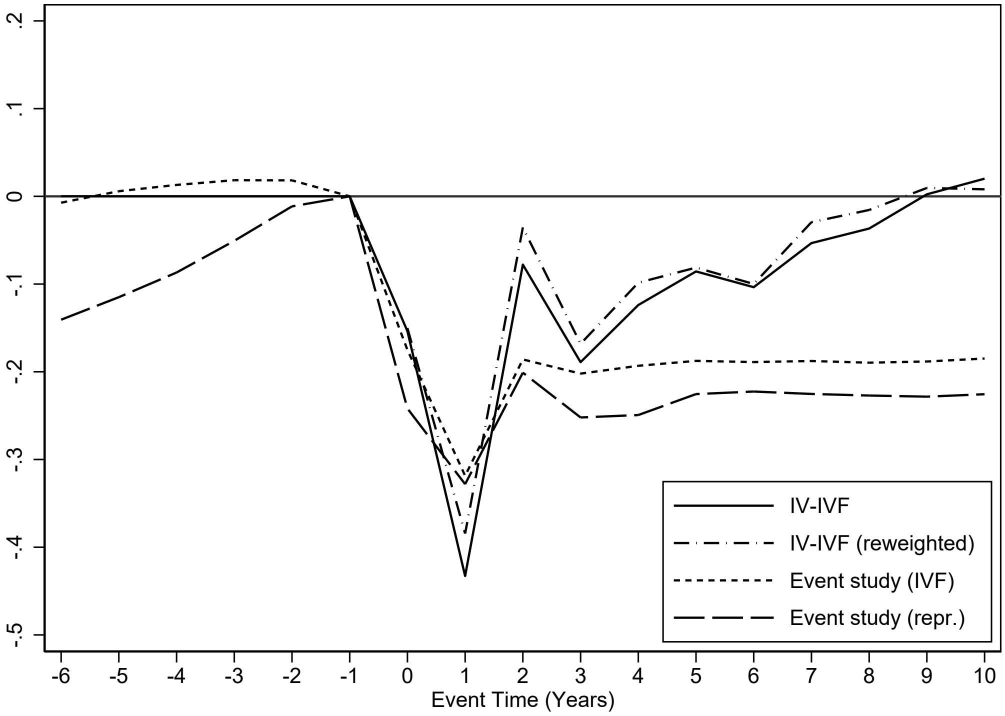

*Notes:*The figure plots the annual earnings of women and ther partners (at the time of treatment) in successfully and unsuccessfully treated couples from 6 years prior to their first IVF treatment up to 25 years after.

Figure 2: IVF treatment effects on fertility, earnings, and gender gaps



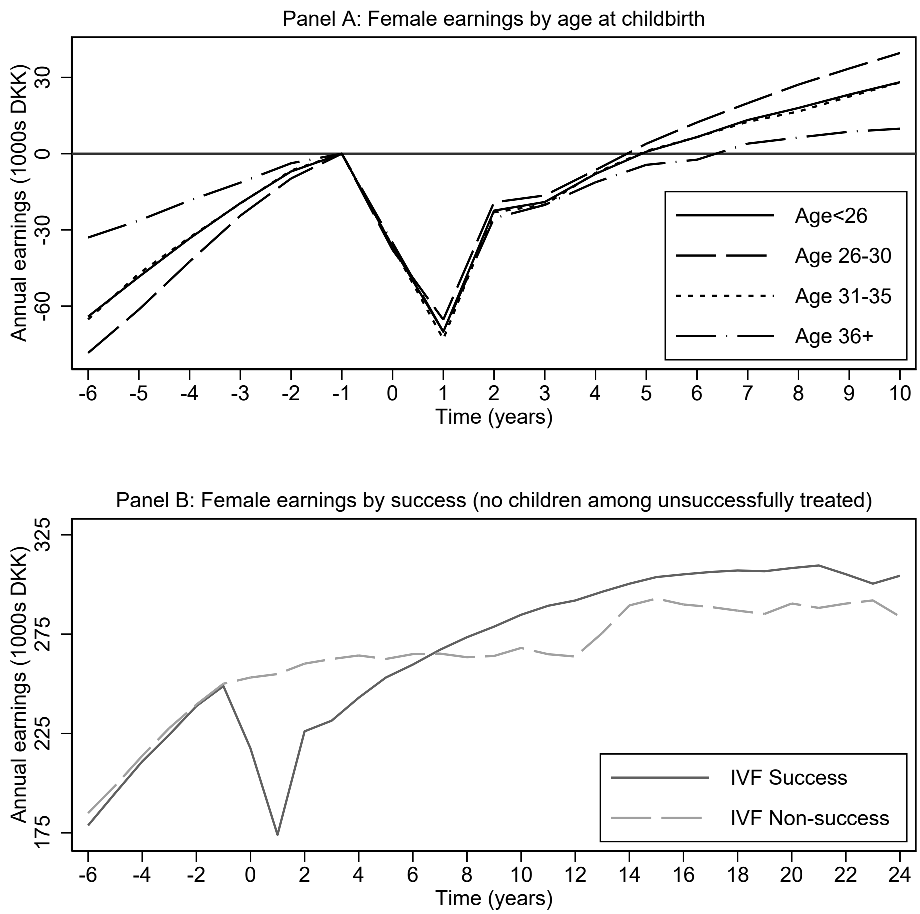

*Notes:* The figures plot the year-by-year estimated impacts of IVF treatment success on fertility (first stage), female earnings, male earnings, and the within-couple earnings gap (reduced form), after controlling for year of treatment, female age at first treatment, and female years of education (in Panels A and B), for year of treatment, male age at first treatment, and male years of education (in Panel C), and for year of treatment, female and male age at first treatment, and female and male years of education (in Panel D); the estimation sample contains IVF-treated couples who had their first IVF treatment between 1995 and 2005; the year of (potential) childbirth is set at 0; the estmated impact in year −1 are normalized to 0. Online Appendix Table 2 (columns 1, 2, 3, and 4) presents corresponding IVF treatment effect estimates based on outcome averages taken over 0-1, 2-5, 6-10,11-15, 16-20, and 21-25 years.

Figure 3: Event study and IVF estimates




*Notes:* The figure plots all the different child penalty estimates for the different samples and strategies: (i) IV-IVF child penalties for all IVF-treated women (solid line); (ii) IV-IVF child penalties for those IVF-treated women who resemble the average representative mother (dashed dotted line); (iii) event-study child penalties for IVF-treated mothers (dotted line); and (iv) event-study child penalties for all mothers (dashed line).

Figure 4: Earnings among unsuccessfully IVF treated women without children




*Notes:* Panel A plots the earnings trajectories of IVF treated women by age at motherhood: below 26, 26-30, 31-35, 36 and older. The annual earnings for all four age groups are normalized to 0 the year before child birth. Panel B plots the earnings trajectories of women with children after a succesfull first IVF treatment (solid line) and women with a failed first treatment and do not have children yet (dashed line). The latter sample includes all women with a failed first treatment up to the point they have a child.

Table 1:
Theoretical First-Stage and Reduced-Form Effects in 3 Period Fertility Model.

| | First-Stage Effect | Reduced-Form Effect |
|---|---|---|
| Estimates | $E(F_t \vert Z=1) - E(F_t \vert Z=0)$ | $E(Y_t \vert Z=1) - E(Y_t \vert Z=0)$ |
| Period 1 | $\beta_c + \beta_a$ ( = 1) | $-\beta_c C_c - \beta_a C_a$ |
| Period 2 | $\beta_c$ | $-\beta_c c_c + \beta_a[C_a - c_a]$ |
| Period 3 | $\beta_c$ | $-\beta_c c_c$ |

*Notes:* This three-period fertility model considers two groups of women who experience different short- and long-run child penalties. The first group consists of compliers: these women remain childless after a first failed IVF attempt. They earn $Y_c$ in the absence of children. After the successful first treatment, represented by indicator $Z$, they experience an earnings loss $C_c$ in period 1 when children are young, and $c_c$ in periods 2 and 3 when children are older. The complier share among IVF treated women is set to $\beta_c$. The second group of women consists of always takers. These women will always end up having children regardless of the treatment outcome at the first IVF attempt. They earn $Y_a$ in the absence of children. After a successful treatment (in either period 1 or 2), they experience an earnings loss $C_a$ when children are young, and $c_a$ when children are older. The share of always takers who experience a successful first treatment is set to $\beta_a$.

# Appendices for Online Publication

Table A1: Descriptive Statistics of Selected Variables.

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| *Pre-treatment:* | Non-success | Success | Repr. sample | (2)-(1) | (2)-(3) |
| Age 1st treatment | 31.966 | 30.906 | 27.832 | -1.061*** | 3.074*** |
| | (4.450) | (3.878) | (4.178) | (0.069) | (0.057) |
| Year 1st treatment | 2000.150 | 2000.293 | 2000.368 | 0.143*** | -0.075* |
| | (3.121) | (3.070) | (3.491) | (0.050) | (0.048) |
| Annual earnings | 244.389 | 243.583 | 204.277 | -0.806 | 39.306*** |
| | (143.428) | (131.772) | (129.544) | (2.269) | (9.93) |
| Schooling | 12.812 | 12.845 | 12.481 | 0.033 | 0.364*** |
| | (2.362) | (2.295) | (2.277) | (0.038) | (0.031) |
| Partner earnings[a] | 323.729 | 318.071 | . | -5.658 | . |
| | (210.542) | (191.286) | . | (3.397) | . |
| Partner schooling[a] | 12.642 | 12.645 | . | 0.003 | . |
| | (2.384) | (2.321) | . | (0.040) | . |
| Positive earnings | 0.908 | 0.922 | . | 0.014*** | . |
| | (0.290) | (0.269) | . | (0.005) | . |
| *Post-treatment:* | | | | | |
| Annual earnings | 258.455 | 244.618 | 158.345 | -13.837 | 86.274*** |
| | (147.351) | (133.174) | (100.285) | (0.232) | (1.395) |
| Partner annual earnings[a] | 370.652 | 367.493 | . | -3.158 | . |
| | (270.022) | (229.904) | . | (0.428) | . |
| Within couple difference | -115.117 | -125.619 | . | 10.502 | . |
| in annual earnings[a] | (275.536) | (238.133) | . | (0.438) | . |
| Observations | 13,176 | 5,370 | 244,435 | | |

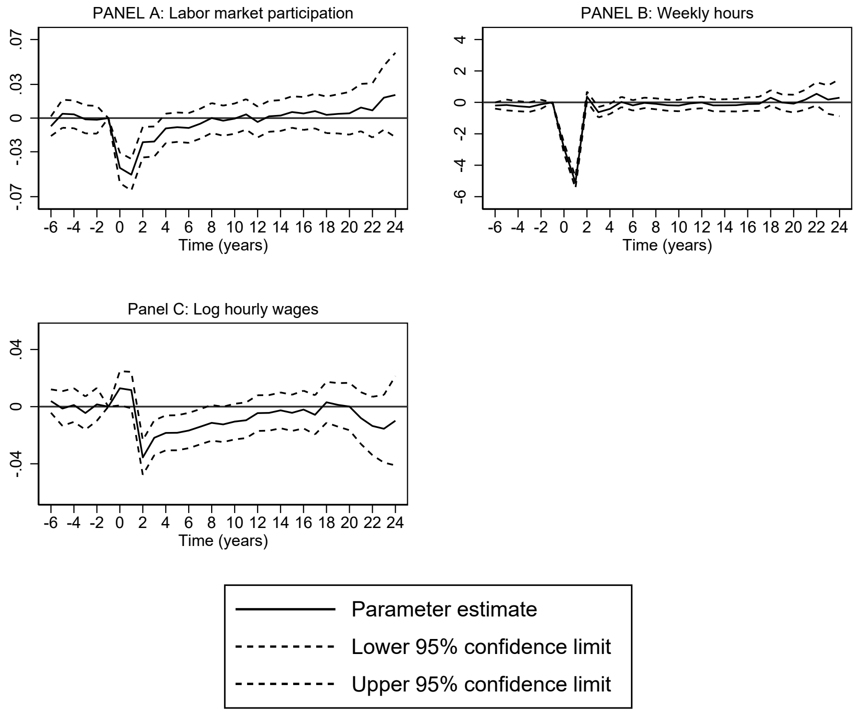
Note: The table shows descriptive statistics for three samples: (1) women having an unsuccessful first IVF treatment, (2) women having a successful first IVF treatment, (3) and a representative sample of Danish women who had their first child born during the study period. Columns (1) to (3) show means with standard deviations in parentheses. Column (4) shows the difference in means between columns 2 and 1 and column (5) shows the corresponding difference between columns 2 and 3. Annual earnings are reported in 1000 DKK and in 2008 Danish Kroner (DKK 100 corresponds to USD 20 as of August 2008). The pre- and post-treatment labor market outcomes represent averages taken over 2 years before and 10 years after treatment, respectively. [a]Partners are the partners at first treatment/birth. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A2: IVF Treatment Effects on Female Labor Outcomes: Results from First-Stage and Reduced-Form Regressions.

| Independent variable | Fertility | Earnings | Partner earnings | Within-couple difference | Within-couple fraction |
|---|---|---|---|---|---|
| | | | *Panel A: Years 0-1* | | |
| *IVF success* | 0.650*** | -46,295*** | -7,159*** | -38,694*** | -0.048*** |
| | (0.003) | (1,425) | (2,488) | (2,853) | (0.003) |
| *N* | 18,546 | 18,546 | 17,374 | 17,373 | 16,989 |
| *F-statistic* | | | | | |
| | | | *Panel B: Years 2-5* | | |
| *IVF success* | 0.286*** | -19,488*** | -5,772** | -4,720 | -0.008** |
| | (0.004) | (1,580) | (2,858) | (3,450) | (0.003) |
| *N* | 18,520 | 18,520 | 17,323 | 17,287 | 16,997 |
| *F-statistic* | | | | | |
| | | | *Panel C: Years 6-10* | | |
| *IVF success* | 0.207*** | -5,507*** | -2,225 | -4,158 | -0.005 |
| | (0.004) | (2,126) | (4,058) | (4,528) | (0.004) |
| *N* | 18,313 | 18,313 | 17,055 | 17,152 | 16,825 |
| *F-statistic* | | | | | |
| | | | *Panel D: Years 11-15* | | |
| *IVF success* | 0.193*** | -2,563 | -2,411 | -0.517 | -0.001 |
| | (0.004) | (2,561) | (4,918) | (5,478) | (0.004) |
| *N* | 18,176 | 18,176 | 16,877 | 17,026 | 16,605 |
| *F-statistic* | | | | | |
| | | | *Panel D: Years 16-20* | | |
| *IVF success* | 0.197*** | 0.627 | -5,114 | 5,862 | -0.001 |
| | (0.004) | (3,234) | (7,362) | (7,888) | (0.005) |
| *N* | 15,299 | 15,299 | 14,077 | 14,334 | 13,810 |
| *F-statistic* | | | | | |
| | | | *Panel E: Years 21-25* | | |
| *IVF success* | 0.233*** | -0.108 | -17,197 | 18,667 | -0.003 |
| | (0.006) | (5,533) | (10,802) | (11,799) | (0.009) |
| *N* | 6,542 | 6,542 | 5,924 | 6,124 | 5,813 |

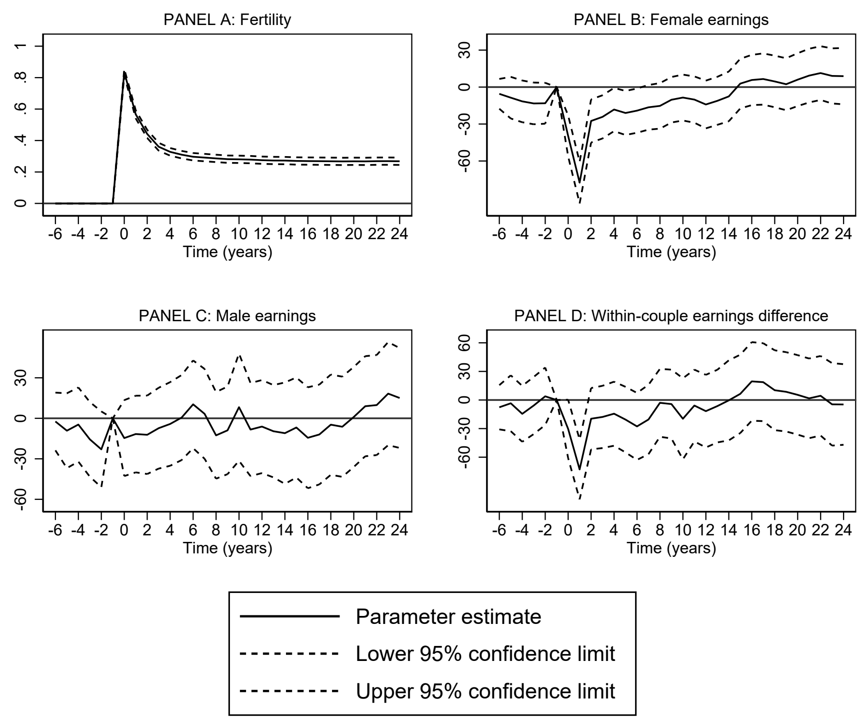
Notes: This table shows the effect of IVF treatment success (0/1) on various outcomes measured at $t$=0-1, 2-5, 6-10, 11-15, 16-20, and 21-25. The outcomes we consider are fertility (column 1), earnings (columns2), partner earnings (column 3), within-couple gender gap in earnings (column 4), and the share of household earnings earned by women (column 5). In column 1, the coefficient represents the effect of IVF success on the probability of having a child during the time period considered. In the reduced form regressions, the coefficient represents the effect of IVF success on the average of the outcome during the time period considered. Time period $t$=0 refers to the year of the (potential) child birth. All regressions control for age at first IVF treatment, year of first IVF treatment, and pre-treatment education and the pre-treatment average of the outcome studied taken over years 1 to 4 before the first IVF treatment. There is one exception. The first-stage regression does not include the pre-treatment average because it is zero by construction. Standard errors are in parentheses; $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.

Figure A1: Other female labor outcomes



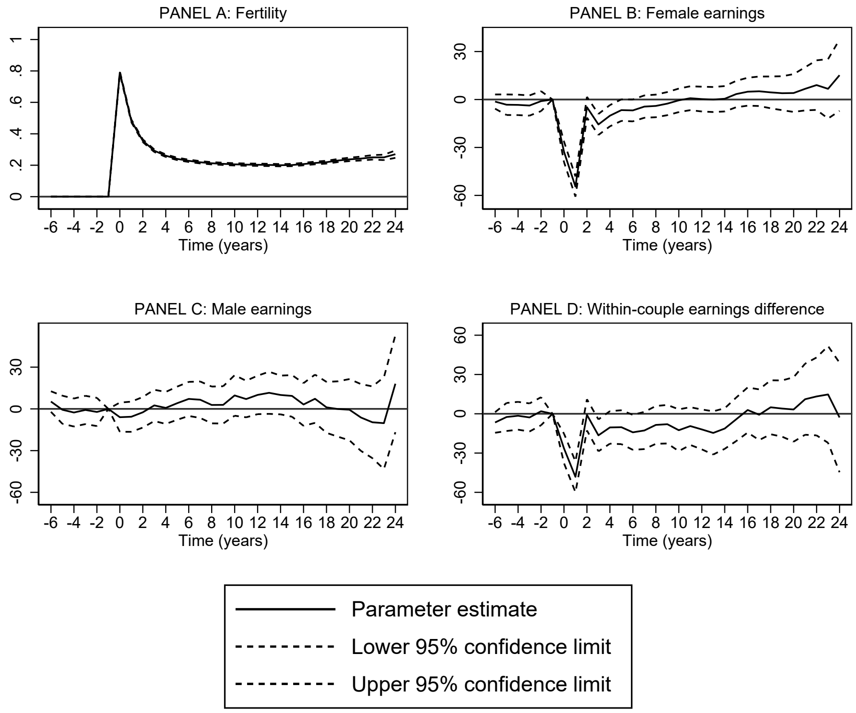

*Notes:* Main results for IVF-treated women revolve around annual earnings. With alternative female labour outcomes (including labor force participation, hours worked, and hourly wages), we plot the year-by-year estimated impacts of IVF treatment success on labor force particiaption (panel A), hours worked (panel B), and hourly wages (panel C), after controlling for year of treatment, female age at first treatment, and female years of education. We refer to the note under Figure 2 for further details. The near zero long-run treatment impact on earnings is also reflected in the other labor market outcomes.

Figure A2: Balanced samples



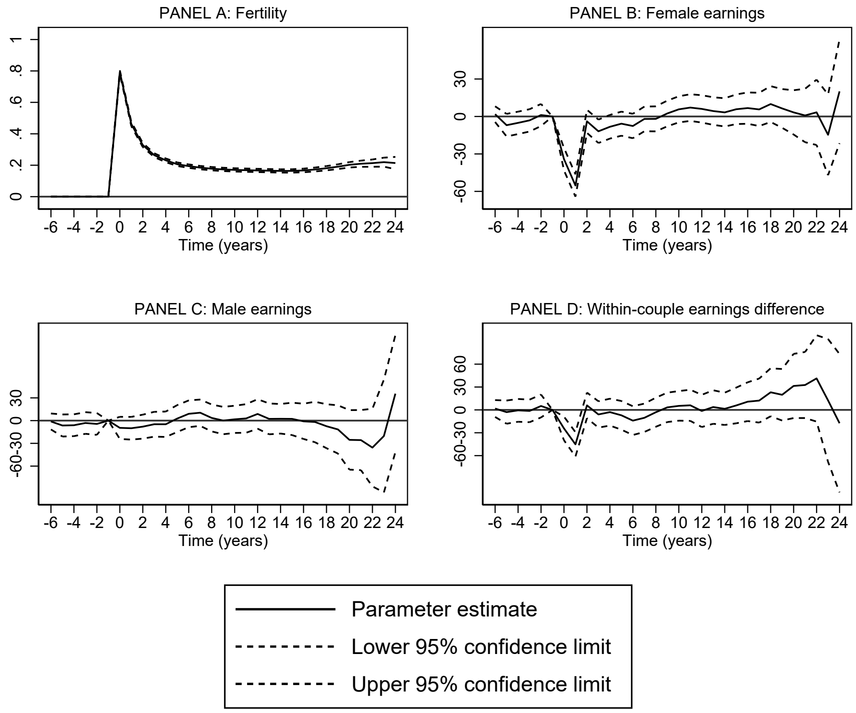

*Notes:* Main results are obtained with an unbalanced sample. By focusing on all first-treated women in 1995 and 1996, we can construct a balanced sample and follow the same women up to 25 years. Based on this balanced sample, we plot the year-by-year estimated impacts of IVF treatment success on fertility (panel A), female earnings (panel B), male earnings (panel C), and the within-couple earnings gap (panel D). We refer to the note under Figure 2 for further details. We conclude that our results using a balanced sample do not really change.

Figure A3: Excluding twins



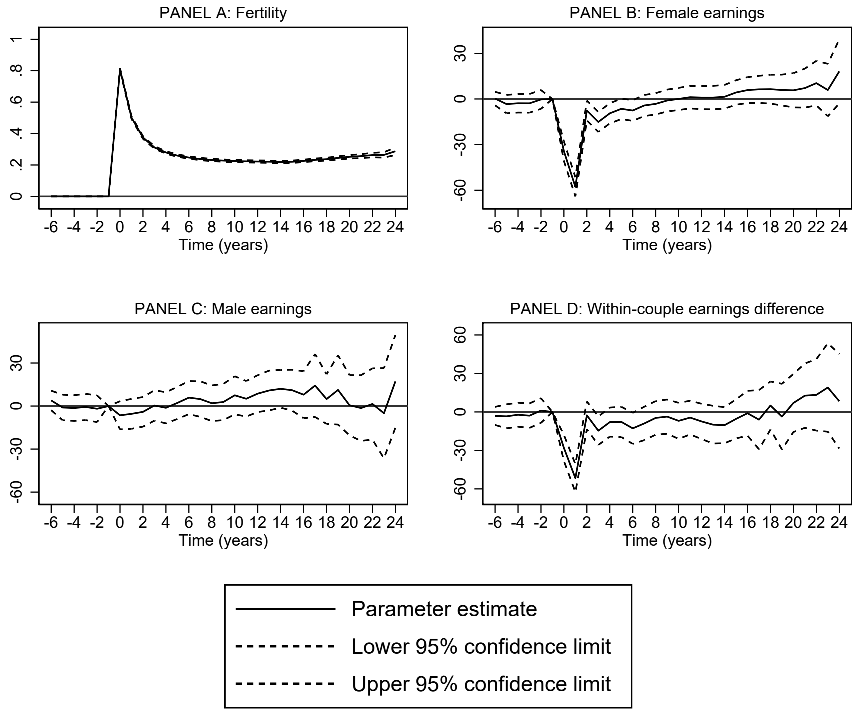

*Notes:* Main results are based on a sample with all births including all IVF-induced twin births. One concern is that mothers may respond very different to twin births (which are common to IVF treatments) than to singleton births. Based on a sample without twin births, we plot the year-by-year estimated impacts of IVF treatment success on fertility (panel A), female earnings (panel B), male earnings (panel C), and the within-couple earnings gap (panel D). We refer to the note under Figure 2 for further details. We conclude that our results do not change when we discard twin births from the sample.

Figure A4: Women whose partners have the infertility problem



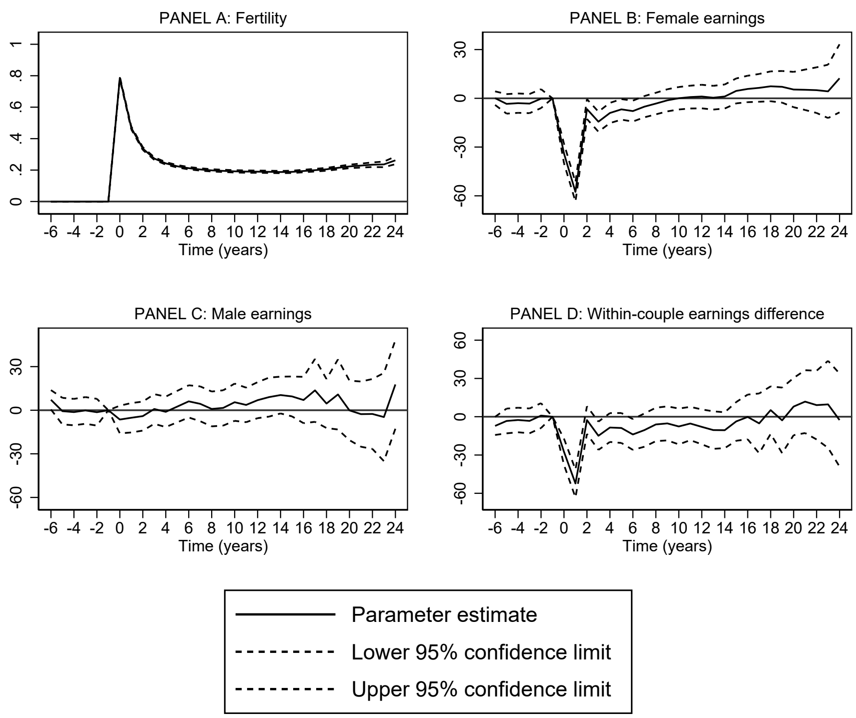

*Notes:* Main results are based on a sample with all IVF-treated women. One concern is that fertility-related health disparities of IVF-treated women may depress their annual earnings. By restricting the sample to IVF-treated women whose fertility problem is on their husband's side, we can estimate child penalties on a sample of healthier women. Based on a sample of healtier women, we plot the year-by-year estimated impacts of IVF treatment success on fertility (panel A), female earnings (panel B), male earnings (panel C), and the within-couple earnings gap (panel D). We refer to the note under Figure 2 for further details. We conclude that our results do not change when estimated on a sample of healthier women.

Figure A5: Without any controls




*Notes:* Main results are based on specifications with varying sets of control variables. With specifications without any control variables, we plot the year-by-year estimated impacts of IVF treatment success on fertility (panel A), female earnings (panel B), male earnings (panel C), and the within-couple earnings gap (panel D). We refer to the note under Figure 2 for further details. Results do not depend on the set of control variables we use.

Figure A6: With all controls




*Notes:* Main results are based on specifications with varying sets of control variables. With the same specifications for all outcomes (controlling for year of treatment, female and male age at first treatment, and female and male years of education), we plot the year-by-year estimated impacts of IVF treatment success on fertility (panel A), female earnings (panel B), male earnings (panel C), and the within-couple earnings gap (panel D). We refer to the note under Figure 2 for further details. Results do not depend on the set of control variables we use.