Health and Labor Market Impacts of IVF Technology: Evidence from a Swedish Policy Mandate

Sonia Bhalotra,[*] Damian Clarke,[†] Hanna Mühlrad,[‡] and Mårten Palme[§]

March 5, 2020

Abstract

A reform mandating that single embryo transfer (SET) replace the default of multiple embryo transfer in in-vitro fertilization (IVF) was implemented in Sweden in 2003. Using linked register data for 1998-2007, we document that the SET reform was associated with a precipitous drop in the share of multiple births of 63%, and a substantial narrowing of baseline differences between IVF and non-IVF births across three relevant domains. For first births, we estimate this convergence to be 58% for child health, 36% for maternal health and 96% for women's income in the three years after birth. Estimates for all births are slightly smaller.

Keywords: IVF, single embryo transfer, fertility, maternal health, neonatal health, career penalty, human capital formation
JEL Codes: J13, I11, I12, I38, J24.

[*]Department of Economics and ISER, University of Essex and IZA

[†]Department of Economics, Universidad de Santiago de Chile and IZA

[‡]Karolinska Institute and Research Institute of Industrial Economics (IFN)

[§]Department of Economics, Stockholm University

[*]We are very grateful to Marianne Bitler, Petter Lundborg, Arizo Karimi, Andreea Mitrut, Randi Hjalmarsson, Mircea Trandafir, Sverker Ek, and seminar audiences at the 3rd IZA Workshop on Gender and Family Economics, Department of Economics at Lund University, Uppsala University, University of Gothenburg, and University of Essex for useful comments and discussions.

1 Introduction

In-vitro fertilization (IVF) is a landmark innovation within assisted reproductive technologies (ART), assisting involuntary infertility and providing women with the opportunity to postpone childbearing. Similar to the introduction of the pill, the legalization of abortion and the availability of long-acting reversible contraceptives (Bailey and Lindo, 2017), IVF has contributed to the economic liberation of women (Abramowitz, 2014, 2017; Kroeger and La Mattina, 2017; Machado and Sanz-de Galdeano, 2015; Rainer et al., 2011). Since its advent in the late 1970s, and tracking significant advances in rates of female labor market participation and contraceptive availability, uptake of this technology has increased steadily over time. As of 2012, more than 5 million children have been born as a result of IVF (Zegers-Hochschild et al., 2009), and the share of all births owing to IVF now exceeds 3% in many industrialized countries (de Mouzon et al., 2010).

There are, however, substantial costs associated with IVF. In addition to costs of the procedure, estimated to range from 12,000 to over 18,000 USD per cycle in the US (Bitler, 2008), there are costs arising from adverse pregnancy and birth outcomes (Kalra and Barnhart, 2011; Saldeen and Sundström, 2005; Sazonova et al., 2011). Women conceiving through IVF treatment are more likely to suffer from complications including hypertension, hemorrhage and emergency C-section. Children born of IVF are more likely to be preterm and small for gestational age. This implies potentially large costs of neonatal and maternal health care (Almond et al., 2010). Biomedical scientists have noted associations of premature delivery with higher infant mortality and with adverse effects on neurological development over a longer term (Gelbaya et al., 2010). Economists have documented causal effects of early childhood health on cognitive skills, educational attainment, life expectancy and income (Behrman and Rosenzweig, 2004; Bhalotra et al., 2017; Bharadwaj et al., 2013; Black et al., 2007; Oreopoulos et al., 2008) indicating additional costs in the longer run.

The main reason that IVF is associated with adverse pregnancy and birth outcomes is that IVF births are 10 to 15 times more likely to be multiple births (Kalra and Barnhart, 2011; Karlström and Bergh, 2007), and multiple births are associated with a higher risk of maternal and neonatal health problems (Bergh et al., 1999; Bitler, 2008; Hall, 2003). For instance, between 2004 and 2005, the rate of twin births among IVF pregnancies was 30% in the United States and 21% in Europe, compared with approximately 1.6% among non-IVF pregnancies in these settings (Maheshwari et al., 2011).[1]

IVF births are more likely to be multiple births because IVF often involves multiple embryo transfers to increase the chances of a successful pregnancy. However, following advances in IVF technology, success rates with a single embryo transfer (SET) have more or less converged to success rates obtained with the transfer of multiple embryos (Criniti et al., 2005; Gerris et al., 2001; Karlström and Bergh, 2007; Kutlu et al., 2011; Lukassen et al., 2005; Lundin and Bergh, 2007; Thurin et al., 2004; Vilska et al., 1999).[2] In light of this medical evidence and growing

[1]A number of studies analyze the negative impact of infertility treatment mandates in the United States on birth outcomes, underlining the healthcare costs of the rising share of multiple births associated with IVF (Bitler, 2008; Bitler and Schmidt, 2006, 2012; Buckles, 2013; Bundorf et al., 2007).

[2]Further discussion of the evidence is in the following section.

concerns over the health costs associated with IVF births, the Swedish National Board of Health and Welfare mandated SET as the default IVF procedure in January 2003.[3]

There were no other changes in the IVF treatment procedure with respect to medication, technique or equipment (Saldeen and Sundström, 2005). Under specified conditions associated with having less healthy embryos, certain women were exempt from SET and continued to qualify for a double embryo transfer (DET) but, barring this, there was a high rate of compliance. This is evident in the share of SET births among all IVF births rising from 30% to 70% in the 24 months following SET (see Figure 1a). The SET reform achieved its goal of lowering costs of neonatal care. One study estimates that, at six months following birth, average costs fell from approximately 160,000 to 90,000 Euros (Thurin et al., 2004), and another estimates that reduced maternal and neonatal hospital stays saved 10,000 Euros per birth (Lukassen et al., 2005).

In this study we document causal impacts of the SET reform for IVF treatments on the twin birth rate and on a rich set of child and maternal health characteristics as well as on maternal income after birth, using linked administrative data.[4] We linked individual mothers and children in birth registers to hospitalisation registers and income files, to create longitudinal data for 1998-2007, a window around the 2003 reform. We first show structural breaks (i.e. a significant single difference) in time-plots of the outcome data for IVF births. We then proceed to incorporate non-IVF births in a double difference approach that allows us to control for aggregate trends.[5]

We now summarise our findings. We find that the SET reform led to a precipitous fall in the rate of twin births from over 30% of all IVF births to 10%. Note that this sharp change was in a direction opposite to the underlying trend of an increase in twin birth rates, which tends to emerge from generalised improvements in the reproductive health of mothers (Bhalotra and Clarke, 2019a,b). The gap in twin birth rates between IVF and non-IVF users narrowed by 66% for first births and 63% for all births. In line with this, after SET, outcomes of IVF-users converged toward those of non-IVF users. The gap between IVF and non-IVF births in an index of child health narrowed by 58% (53%) for first (all) births, by 36% (19%, and imprecise) for an index of maternal health and by 96% (85%) for maternal income in the three years after birth.[6]

The results are robust to including mother fixed effects, allowing differential trends for the outcomes of IVF and non-IVF mothers, and to allowing these trends to vary by a rich set of baseline characteristics of the mother. This undermines the potential concern that our results are driven by endogenous changes in the composition of mothers adopting IVF.

We find significant improvements in ten of thirteen indicators of child health, including gestational age, birth weight, birth length, head circumference, APGAR score, breech presentation,

[3]In the US in 2012, it was estimated that the average cost of a singleton birth was $27,000, while twin and triplet births cost $115,000 and $435,000 (Lemos et al., 2013).

[4]We avoid contamination with individual variation in the risk of multiple birth created, for instance, by differences in baseline maternal health by leveraging only the variation in risk induced by the SET reform.

[5]This also provides a natural benchmark for analysis of effect sizes: we express all improvements in outcomes of IVF births in terms of the fraction of the pre-reform gap between IVF and non-IVF births that was closed by mandating SET. We consistently present results for first-births (71% of IVF births) and for all IVF births.

[6]We do not expect full convergence. First, we noted that only 70% of IVF births used SET after the reform. Second, the medical literature has noted that singleton IVF births are less healthy than singletons non-IVF births (Pinborg et al., 2013; Sazonova et al., 2011) and we confirm this using our data.

hospitalisation in the first year of life, and infant and under-5 mortality. The three indicators of child health that are not significantly different after the SET reform are the probability that the IVF birth is male, exhibits fetal malformation, or is hospitalised at age 1-4. The impacts of SET are large. For instance, birth weight increases by 175 grams in the sample of all births and by 186 grams among first births, closing the gap relative to non-IVF births by 57 and 63% respectively. For the two most commonly used indicators in the literature, birth weight and gestational age, we investigated impacts across the distribution, and found gains across the distribution. Improvements in maternal health are driven by a decline post-SET in the risk of emergency C-section of 2 (all births) and 3 (first births) percentage points, which narrows the gap with respect to non-IVF births by 42 and 60% respectively. These results have implications for healthcare costs associated with IVF births, and also potentially for the long term health and productivity of post-SET cohorts.

The SET reform also had immediate productivity effects. We find that the labor income of IVF mothers averaged over the three year period following birth is higher for post-SET births by 7% for women having their first birth and 8% for all women. These increases not only close the pre-SET gap in labor income between IVF and non-IVF women, they reverse the gap. This is not implausible as women who use IVF tend to be more educated.[7] We additionally find statistically significant changes in benefits (discussed later), though these are somewhat more mechanical.

Overall, this paper provides the first comprehensive analysis of the benefits of mandating single embryo transfer in IVF procedures, using administrative data that contain unusually rich data on child and maternal health and that track women's earnings for three years following birth. It shows that mandating SET not only significantly lowered the risk of neonatal and maternal health complications but it also lowered the career costs of fertility among IVF-using women.

A number of biomedical studies have shown that the SET mandate was associated with a sharp drop in twin births (Karlström and Bergh, 2007; Lundin and Bergh, 2007; Saldeen and Sundström, 2005; Sazonova et al., 2011; Thurin et al., 2004). There is also some evidence in this literature that specific measures of neonatal and maternal health are better for IVF pregnancies that result in a singleton rather than a twin birth (Sazonova et al., 2013). We use a much larger sample than most of this literature, covering the entire population of more than 0.93 million births in Sweden in the analysis period. We use different methods, in particular, differencing with respect to non-IVF births, adjusting for trends, and analysing endogenous compositional change. We analyse a wider range of outcomes, for instance we study not only indicators of health at birth but also longer term health outcomes for children including the risk of hospitalization up until age 4 and infant and under-5 mortality. Our largest substantive contribution to this literature lies in studying impacts of the SET reform on the earnings of mothers in the years after birth.

We are unaware of studies in the economics literature that evaluate a SET reform. However our paper relates to a literature showing that fertility leads to a sustained loss in earnings for women (Adda et al., 2017; Kleven et al., 2019; Lundborg et al., 2014).[8] In our setting, the post-SET in-

[7]We check that there is no significant change in the educational composition of IVF-users after SET.

[8]Lundborg et al. (2014) uses a sample of IVF women in Denmark, exploiting idiosyncratic individual variation in IVF success rates to identify the impact of fertility on women's earnings. They compare women who have a successful IVF pregnancy resulting in one or two births with women who are unsuccessful with IVF and thus have no birth. In contrast, we use a sample of women who succeed with IVF, effectively comparing women who have one birth with

crease in earnings of IVF-using women may in fact result from the fact that they or their children are in better health, see Bhalotra et al. (2018); Bloom et al. (2015) for evidence of these channels determining women's labour force participation. Constrained by the fact that we effectively have one instrument (the reform) rather than multiple instruments, we present a descriptive decomposition exercise that indicates that the higher earnings of post-SET IVF-using women emerge from lower fertility (lower rates of twinning) rather than from improved health.

Our findings have important implications for other countries considering policy reform. Following the lead of Sweden, other countries including Belgium and Turkey have mandated SET, but double or multiple embryo transfers are still prevalent in most other countries including the US and the UK. For example, only 10% of all embryo transfers were single transfers in 2008 in the US (Practice Committee of the Society for Assisted Reproductive Technology and Practice Committee of the American Society for Reproductive Medicine and others, 2012). Elective adoption of SET tends to be inhibited for two reasons. First, some women may have a preference for twin births. Second, pregnancy success rates may be perceived to be lower with SET, leading to multiple attempts at IVF. In environments in which families privately bear a large share of the costs of IVF, they may elect for DET even when they do not have a preference for twins (Hamilton et al., 2018). Thus, countries like the US with relatively restrictive public coverage for IVF have a harder time implementing SET (Karlström and Bergh, 2007; Pinckney-Clark et al., 2016).

The rest of this paper is structured as follows. Section 2 describes IVF and the SET reform in Sweden, as well as the data used here. Sections 3 discusses the empirical strategy. Section 4 presents the results and Section 5 concludes.

2 Background

2.1 Pregnancy success rates with IVF treatments

As discussed in the previous section, IVF treatments have typically involved multiple embryo transfers on the premise that this raises the odds of a successful pregnancy. Before the reform that we analyse, most IVF clinics in Sweden implemented a voluntary reduction in the number of embryos routinely transferred, from three to two in 1993. This resulted in the virtual elimination of the conception of triplets by IVF, while the pregnancy rate and the live-birth rate remained essentially unaffected at 35 percent and approximately 25 percent per transfer, respectively (Thurin et al., 2004). Then observational studies set in clinics in Finland and Belgium respectively, involving approximately 1000 patients each, demonstrated that pregnancy success rates were not significantly lower even when the number of embryos transferred was reduced from two to one (Gerris et al., 2001; Vilska et al., 1999).

A major randomised control trial supporting the SET mandate was published a year after the reform– a study involving 661 participants in clinics spread across Sweden, Norway and Denmark (Thurin et al., 2004).[9] Subsequent evidence emerging from trials and observational studies re-

those who have two.

[9]This study showed that the success of IVF was maintained with SET under certain circumstances, namely when the woman was below 36 years and had at least two embryos of good-quality. They found that the cumulative rate of

inforced the broad conclusions of small and insignificant differences in pregnancy success for a broad class of women (Criniti et al., 2005; Karlström and Bergh, 2007; Kutlu et al., 2011; Lukassen et al., 2005; Lundin and Bergh, 2007).

2.2 IVF treatments in Sweden- access and eligibility

All permanent residents in Sweden have access to heavily subsidized health care offered by both private and public health care providers. For most medical services, there is a small fee until the patient reaches the maximum amount of 1100 SEK (approximately 110 USD) annually. Health care services have usually no additional costs. Health care is mainly funded by tax revenues and only 2% of residents have private health insurance (Anell, 2008). Sweden has adopted a series of 'family friendly' policies, including access to 16 months of paid maternal and paternal leave, paid parental leave for the long term care of a sick child, and heavily subsidized child care. IVF procedures are regulated under the law on genetic integrity.[10] The Swedish Association of Local Authorities and Regions provides guidelines for eligibility for IVF treatments. The local health care provider is responsible for adherence, and enforcement is often not strict (SKL, 2016).

Eligibility criteria. First, the couple should be in a stable union, either legally married or co-habitating for at least two years, although since 2016 single women are also allowed to access publicly funded IVF treatment, and lesbian couples are allowed access since 2005. Second, the woman should have no previous children, either biological or adopted. IVF is available for second and higher order births but this is not publicly funded. Third, a medical assessment of the woman should be completed to confirm that her body mass index (BMI) is within the normal range, that there is no evidence of risky behavior such as smoking and use of alcohol and other drugs/narcotics. Other mental and physical illness and disability are also considered before offering treatment. The suggested maternal age for starting the first treatment is below 40 and any remaining embryos/egg cells should be transferred before age 45. The age of the man should lie between 25 and 56 years. Age restrictions are county-specific, for example, the maximum age of the mother in Örebro county is 43 while in Norrbotten county it is 37 (Alm, 2010). Fourth, three rounds of treatment (follicle aspiration) are offered to each couple, and any remaining embryos and eggs of good quality are frozen.

As discussed, on January 1 2003, the Swedish National Board of Health and Welfare mandated SET as the default IVF procedure, but it allowed exceptions for women with a low perceived risk of twinning. In particular, women with low embryo quality, those aged above 38 years and/or those women with more than three previously failed IVF cycles were still allowed DET, provided that they were informed about the potential risks for the mother and child (Saldeen and Sundström,

live births was not significantly different between elective SET (38.8%) and DET (42.9%), this being the probability of at least one live birth following transfer of one fresh embryo (under SET), and if needed, a subsequent transfer of a frozen embryo.

[10] In Swedish: "Lag (2006:351) om genetisk integritet m.m.". Other aspects relating to IVF treatment such as establishing parenthood and defining and protecting patient rights are regulated by other laws, including the Children and Parents' Code (Föräldrabalk (1949:381)) and the Health and Medical Services Act (Hälso-och sjukvårdslag (1982:763)). IVF using donated gametes is only permitted in publicly funded university hospitals under the law (2013:1147). For donated gametes an extraordinary assessment is required according to law (2016:18), with requirements similar to an adoption process.

2005). We highlighted in the previous section that, at the time that SET was implemented, there were no other changes in the IVF treatment procedure with respect to medication, technique or equipment (Saldeen and Sundström, 2005).[11]

2.3 Data

The analysis sample is constructed by linking several administrative data sets. The population consist of all births in Sweden during 1998-2007 identified via the Medical Birth Registry provided by the National Board of Health and Welfare, which covers approximately 99% of all births in Sweden since 1973. This register sources its information from prenatal care units, maternity clinics and neonatal care units. It collects information on fertility treatments including standard IVF, Intra cytoplasmic sperm injection (ICSI), surgical procedures and ovarian stimulation, conditional upon the treatment resulting in a successful pregnancy delivered after week 22. For this analysis we focus on births that are the product of IVF procedures including standard IVF and IVF with ICSI. We merge in individual data from the National Patient Registry on the number of nights spent in hospital by the mother and child and information on mortality from the Cause of Death Registry, both of which are available from the National Board of Health and Welfare. Administrative data on income and educational attainment of mothers is obtained from the Social Insurance Agency and the Swedish Agency for Innovative Systems, provided by Statistics Sweden.

During the analysis period of 1998-2007 (60 months before and after the SET reform), the registry recorded 21,783 IVF births and 916,110 non-IVF births. Thus in these data the share of IVF births is 2.3%. Of all IVF births, 71% are first births.[12] We remove triplet and higher order births (516 births in all). Given potential within-pregnancy variation in some outcomes (for example child health outcomes in the case of twin births), the unit of observation is births and as such twins will represent two outcomes.

We have demographic data for the mother and child including gender, parity and singleton/multiple birth status of the child and the education and age of the mother.[13] For the mother we also observe weight, height, chronic diseases, tobacco consumption and prenatal conditions and treatments such as the use of fetal diagnosis service and pregnancy complications (diagnosis and procedures).

We classify the outcomes of interest into three domains as follows. **Child health**. APGAR score, birth weight, birth length, head circumference, fetal malformation, breech presentation, gestational age, the probability the child is male, nights hospitalized during the first year of life and

[11] There is one exception. In January 2003, coincident with the SET reform, there was a change in regulation (*Socialstyrelsens föreskrifter och allmänna råd om assisterad befruktning* SOSFS 2002:13) that allowed donated eggs or sperm to be used in IVF treatments, although subject to an extensive assessment of the couple's medical, psychological and socio-economic characteristics, similar to those in an adoption process (Socialstyrelsen, 2016). Also the amendment allowing donated gametes was restricted to publicly funded university hospitals. In 2002, only 19 IVF cycles using donated egg cells were attempted resulting in 6 live births (Socialstyrelsen, 2006). While the number of IVF cases with donated eggs cells has increased (from 19 cycles in 2003 to 401 cycles in 2010, resulting in 86 live births), the share of IVF births using donated eggs cells is only 2% of all IVF births (Socialstyrelsen, 2013).

[12] This figure of 71% refers to births regardless of multiplicity, ie are at the level of the mother.

[13] We will control for the highest level of education of the woman, a categorical measure from level 1-7. Level 1 is primary education less than 9 years, level 2 is primary education of 9 years, level 3 is 2 or fewer years of secondary education, level 4 is 3 years of secondary education, level 5 is fewer than 3 years of tertiary education, level 6 is 3 or more years of tertiary education and level 7 is graduate-level studies.

during years 1-4, infant mortality and under-5 mortality. For some indicators we create more than one outcome, for example, low birth weight and birth weight. For birth weight and gestational age, we also present results for medically relevant cutoffs, and plot effects across the distribution.[14]

Maternal health. Use of emergency C-section, maternal sepsis, postpartum hemorrhage, hypertension in pregnancy, and hospital readmission the first year after delivery.[15]

Maternal income. Income from gainful employment, parental benefits and sickness benefits averaged over the three years after birth are expressed in real terms using the 2016 consumer price index.[16] The data do not provide employment or hours of work, only earnings. Only 8% of all women have zero earnings within 3 years of giving birth. We present estimates for earnings (and benefits) in Swedish kroner including the zeroes and separate estimates excluding them and modelling the probability of shifting from zero to nonzero earnings after SET. We also provide estimates of the impact of SET across the earnings distribution.

2.4 Descriptive Statistics

First we compare pre-SET or baseline characteristics of IVF vs non-IVF users averaging the data for 1998-2002 (Table 1). The rate of twin birth in the IVF sample is 30.3%, compared with 2.5% in the non-IVF sample. Women using IVF are older, taller, heavier and less likely to have smoked during and before pregnancy (in line with Bhalotra and Clarke (2019b)—note that the sample only consists of the IVF users who succeeded in having a birth).[17] They have higher education and earnings. Despite these risk-reducing characteristics, they are more likely to suffer a range of birth-related complications. Their children have worse indicators of health at birth.

Since the SET reform led to a drop in twin births among IVF users, we move on to compare IVF and non-IVF users in the pre-SET period conditional upon whether the birth was a twin or a singleton (Table 2). On most indicators of child health, IVF-twins are not significantly different from non-IVF twins.[18] However, IVF singletons have significantly worse indicators than non-IVF singletons. This may reflect, inter alia, that singletons born after an IVF procedure were conceived as twins, with only one surviving to birth. It is important here because it tells us that we should not expect complete convergence of IVF outcomes towards non-IVF outcomes after the SET reform.[19]

[14] APGAR, measured 5 minutes after birth, stands for "appearance, pulse, grimace, activity, respiration" and is a five-criterion evaluation method, indicating the general health condition of the newborn baby 1, 5 and 10 minutes after the delivery.

[15] Postpartum hemorrhage (severe blood loss) and maternal sepsis (infection) are regarded as severe maternal complications. Sepsis is defined as "infection of the genital tract occurring at any time between the rupture of membranes or labor, and the 42nd day postpartum, of which two or more of the following are present: pelvic pain, fever 38.5 C or more, abnormal vaginal discharge, abnormal smell of discharge, and delay in the rate of reduction of size of uterus (less than 2 cm a day during the first 8 days)" by the WHO (Bamfo, 2013).

[16] These variables are measured (respectively) as: total annual gross earnings in cash and net income from active business; total annual income from parental leave including income from parental allowance, temporary parental leave and child care allowance; total annual income caused by illness, injury and/or rehabilitation including a sick pay period of 14 days.

[17] Fitted plots of the age distributions in the two samples are provided as Appendix Figure A1.

[18] One explanation of this is that non-IVF twins are more likely than IVF twins to be positively selected on maternal health and education (Bhalotra and Clarke, 2019b).

[19] The singleton sample is much larger than the twin sample and this will contribute to more precise estimation of differences in the singleton sample. However we note that the medical literature (Pinborg et al., 2013; Sazonova et al.,

The Table shows that many maternal characteristics are significantly different between IVF and non-IVF mothers irrespective of whether the birth is a twin. Notice that while mother's earnings after birth are higher among IVF mothers (in line with their higher education) in both samples, they are clearly higher for IVF mothers bearing a singleton than for IVF mothers bearing a twin. This is formalised in our estimates of the impact of the SET reform.

2.5 Post SET trend breaks in outcomes

A sharp increase in the share of IVF treatments using SET, from 30% to 70% within 24 months, is displayed in Figure 1a. The pregnancy success rate among IVF users was maintained at about one-quarter (Karlström and Bergh, 2007), see Figure 2a. There is no evidence that the reform either encouraged or discouraged uptake of IVF treatments, see Figure 2b, which shows that the number of IVF treatments performed is smooth around the cut-off.[20] This limits concern about endogenous shifts in the composition of mothers using IVF after SET but we nevertheless investigate this later. As we discuss in sections 3 and 4, certain areas of Sweden pre-empted the SET reform given the arrival of medical evidence from the SET studies as early as 2001. Thus, later in the paper, as well as estimating the ITT of the SET policy implementation in 2003, we will also discuss specifications removing these two earlier years, and removing a single county which adopted SET prior to 2003.

The sharp drop in the share of twin births among IVF conceptions from 30% to 13% is evident in Figure 3, which also depicts a stable path of the share of twins among non-IVF births. As a prelude to the analysis, we show unconditional outcome data plots in Appendix Figures A3, A4 and A5. Most of our outcome indicators for child and maternal health and for maternal income show a sharp improvement after the 2003 reform.

3 Empirical strategy

We exploit exogeneity in the timing of the SET reform for IVF-users, differencing with respect to non-IVF users to allow for common trends. The estimated equation is:

$$Y_{it} = \alpha + \beta_1 (PostSET \times IVF)_{it} + \beta_2 IVF_i + \boldsymbol{X}_{it}\delta + \alpha_c + \pi_t + \varepsilon_{it}, \tag{1}$$

where the dependent variable Y_{it} refers to a birth or maternal outcome for birth i in year t, and IVF_i refers to the IVF status of each birth (1 if IVF was used, or 0 otherwise). $PostSET$ is a binary variable based on estimated date of *conception*: all births estimated to have been conceived after January 1 2003 are assigned as $PostSET = 1$. Rather than include the main effect $PostSET$ in the regression, we include a series of year fixed effects π_t to flexibly control for all relevant time varying unobservables. County-specific fixed effects α_c capture time-invariant geographical variation in the outcomes. The control variables $\boldsymbol{X}$ are maternal age and pregnancy order fixed effects, maternal height and weight before pregnancy, nationality (a binary variable for having

2011) has also noted that IVF singletons have worse birth indicators.

[20] Figure A2 shows that there is a secular trend in the proportion of IVF births and also in the share of twin births in all (IVF and non-IVF) births. Trends in the proportions of each type of ART procedure presented in Figure A2 show that IVF is the only ART procedure exhibiting a trend.

been born in Sweden or not), the mother's educational level, an indicator for the mother having smoked during the first trimester of pregnancy and the value of her sickness benefits and labor income averaged over the three years prior to birth. Standard errors are clustered by mother and we use the ordinary least squares estimator.

The parameter of interest is β_1, capturing the change in outcomes for IVF births relative to non-IVF births after 1 January 2003. It is a reduced form intent to treat parameter. Our estimates allow us to capture not only the SET-led improvements in IVF outcomes (β_1) but also the extent to which SET led to a convergence of IVF birth outcomes with non-IVF birth outcomes (β_1/β_2). The data do not identify at the individual level which women used SET and which were exempt and allowed to use DET, though we know from aggregated figures provided by the government that about 70% of post-reform IVF births used SET. On account of this fact of imperfect compliance, together with the fact that IVF singletons have worse health than singletons born following an unassisted conception (shown in preceding section), we do not expect absolute convergence of IVF to non-IVF outcomes.

Identifying assumption. This is that, in the absence of the SET reform, outcomes associated with IVF and non-IVF births would have followed similar trends over time. In order to test the plausibility of this assumption we estimate an event study, interacting the IVF indicator with an indicator for each year before and after the reform date. The specification we estimate is:

$$Y_{it} = \alpha + \sum_{k \in \ell} \gamma_k (IVF_i \times \mathbb{I}\{Year_t = SET + k\}) + \beta IVF_i + \boldsymbol{X}_{it}\delta + \alpha_c + \pi_t + \nu_{it}, \tag{2}$$

where $\ell = \{-4, -3, -2, 0, \ldots, 4\}$ and the year before the SET reform, 2002, is omitted as a base category. A test of the equality of the lagged coefficients will indicate if IVF and non-IVF outcomes exhibited differential pre-trends.

To further investigate the concern that differential pre-trends may bias our estimates, we include group-specific trends in the model. In other words, we interact the IVF indicator with a linear trend. We estimate two variations of this, a linear trend over the entire period (which we shall call a *global trend*) and an IVF-status-specific linear trend that is allowed to be different in the pre vs the post-SET period (which we shall call a *split trend*).

Endogenous changes in sample composition. We saw, in Figure 2b, that there is no structural break in the trend in women using IVF coincident with SET. However, it remains possible that women who use IVF after the SET reform are different from women using IVF before SET. This raises the potential concern that endogenous changes in composition drive our results, in particular, that post-SET IVF users have more favourable characteristics that translate into lower chances of twinning, better child and maternal health and higher maternal income after birth.

First, we would argue that this possibility is undermined a priori because more favourable maternal characteristics would lead to higher twinning alongside better health and earnings. We nevertheless conduct the following checks. We present tests of balance of characteristics of women using IVF before vs after the reform and account for any differences by introducing controls for baseline characteristics of the mother in the model, denoted $\boldsymbol{X}$. Using the specifications with and without these controls for mother characteristics, we compute the Altonji-Taber statistic to estimate the chances that accounting for unobservables would change the coefficients of interest.

We estimate an additional specification that allows for unobserved mother-level heterogeneity using the roughly 50% of women who use IVF and have more than one birth, to estimate the model conditional on mother fixed effects. In another check, we report results from adding to the baseline model IVF-status specific trends interacted with maternal characteristics. Finally, we investigate heterogeneity in impacts of the reform by mother characteristics including parity, education, age, BMI. In the case of parity, we systematically estimate all results for all women and then again for first-time mothers (44% of the full sample), and first births (71% of the IVF sample).

Multiple hypothesis testing. We have many indicators of the outcomes. We are thus faced with a problem of multiple-inference and risk over-rejecting null-hypotheses (i.e. an inflated rate of Type I errors). We address this issue using two different approaches. First, we create summary indices for child health, maternal health and maternal income, which decreases the number of hypotheses tested. The indices are constructed following Anderson (2008), and they attach more weight to variables which contribute more independent variation to the aggregate.[21]

Second, we adjust p-values by controlling for the false discovery rate (the proportion of Type I errors in all significant findings) among all variables examined, using a step-up procedure described by Benjamini and Hochberg (1995). This method has the advantage of greater power compared to other approaches but at the cost of allowing for the false rejection of null-hypothesis (Anderson, 2008). We also report the considerably more demanding Bonferroni (1935) corrected p-values, which control for the Family Wise Error Rate, and thus set the size of each test to avoid falsely rejecting *any* hypothesis.

Timing and measurement error. As SET was implemented two years earlier in the county of Skane, due to a local initiative, we re-estimate the model excluding this county. We also re-estimate the model excluding the two years during which we see a gradual increase in the share of SET births among IVF mothers, so that identification comes from a sharp discontinuity in this share. This also addresses a different problem. We use the date of conception to define pre vs post SET births but, as is common, conception date is not directly available and is estimated with error. We compute it by subtracting the length of the gestational period in days from exact date of birth, analogous to Currie and Schwandt (2013). However, date of birth is not available in our data set, so we proxy this with the date of discharge from the maternity unit. So, for births (/mothers) that have longer hospital stays after birth, we will estimate a later date of conception than the true date. This means that we will tend to classify as post-SET (treated) some births that are in fact pre-SET (untreated) and these will systematically be the more complicated births. This implies we will tend to under-estimate the true (positive) impact of SET.

[21] We first ensure that all variables are consistently measured so that more positive values imply a positive change, for example, when considering the variables birth weight and premature, prematurity is multiplied by -1 so that both birth weight and "not premature" refer to positive health measures at birth. Then all variables are standardized by subtracting the mean and dividing it by the standard deviation of the variable in the control group. Finally, indices are created using a weighted average of the standardized variables of interest. Each variable is weighted by the inverse of the covariance matrix among the full set of variables and as a result those contributing the most linearly independent information receive a higher weight in the index.

4 Results

4.1 Twin birth rate in IVF pregnancies

Table 3 presents the impact of the SET reform on the likelihood that an IVF birth is a twin as opposed to a singleton birth (row 1). For the full sample (columns 1-2) we estimate a reduction in the share of twins of 16.8 to 17.3 percentage points (pp), depending on whether or not we control for mother characteristics. Estimates for first-time mothers (columns 3-4) are very similar at 17.7 pp, and not sensitive to controls. The main effect of IVF (row 2) shows that IVF births in the pre-reform period were 27 pp more likely to be twins. Using this to scale the estimates in row 1, we conclude that the SET reform narrowed the gap in twinning between IVF and non-IVF births by about 63%.

Group-specific trends. To account for omitted trends specific to the outcomes of IVF-users, we include IVF-specific *split* linear time trends (columns 1 and 3) and IVF-specific (*global*) linear time trends (columns 2 and 4), see Table A1. We continue to see a significant drop in the share of twins in IVF relative to non-IVF births after the reform, but the magnitude drops from about 17 pp to about 13 pp.

Heterogeneity. Estimates of investigate heterogeneity in impacts by baseline characteristics of the mother including parity, age at treatment, education and BMI are in Table 4. The overwhelmingly clear pattern is that mothers of all types respond broadly similarly to the reform. We see a statistically significant reduction in the share of twins among IVF births in every age group other than of women 39 years and older. These age 39-plus results provide a placebo test since these women were likely to be exempt from the SET mandate in most counties, partly because they have a lower probability of twinning (see the coefficient on IVF in Panel A of the table: the probability is 14% compared with about 27% on average). We observe similar sized reform-led effects on twinning at parities 1 and 2 and somewhat smaller effects at parity 3 and higher; similar effects for women of all ages up to 35, and smaller effects for older women, no significant differences by the woman's education, and similar effects across the BMI-range except slightly larger effects for under-weight women. Comparing across columns the ratio of the interaction effect to the main effect, any differences are small, consistent with the mechanism driving the decline in twin births across different types of mothers being biological rather than behavioural.

4.2 Child health outcomes

The main results for the three outcome domains are in Table 5. We identify a significant improvement in the index of child health for IVF births of 0.189 standard deviations (henceforth, SD) post-SET which is similar for first-time mothers and all mothers. Relative to the main effect of IVF, this implies that the SET reform acted to reduce the health gap between IVF and non-IVF children of -0.355 SD by 53%.

Components of the index. The thirteen indicators of child health are separately analysed in Table 6. The three indicators that are not significantly modified by SET although their coefficients move in the expected direction are fetal malformations, the probability that the child is male, and child hospitalization at ages 1-4. For the other ten, we find statistically significant and typically large improvements in IVF birth outcomes post-SET. For example, average birth weight increased by 175 grams, closing the gap between IVF and non-IVF babies by 57%, gestational age increased by more than half a week on average, closing the gap by 52%. Changes in birth length and head circumference similarly narrowed the IVF–non-IVF differential by close to 50%. Results are similar and slightly larger for first births. To account for multiple hypotheses when examining individual child health components, we correct the p-values, first with a false discovery rate correction and then also using the more demanding Bonferroni correction, these estimates are in Table 6. Our conclusions are entirely robust to these adjustments.

Distribution of effects. For two of the most commonly used indicators of birth outcomes, birth weight and gestational age, we estimate distributional impacts. In Appendix Table A2 we use cut-offs that are often used in the targeting of medical resources (Almond et al., 2010; Bharadwaj et al., 2013). We see that, following the SET reform, the likelihood of IVF babies being born with a weight below 1500 grams (very low birth weight) fell by 1.2 pp, and the likelihood of being born with a weight below 2500 grams (low birth weight) by 6.8 pp. Scaling by pre-reform differences between IVF and non-IVF babies, the proportional impacts are both about 60%. The probability of preterm delivery before weeks 28, 32 and 37 decreased by 0.5 (63%), 1.3 (52%) and 8.3 (53%) pp. Very similar results are found for first-time mothers (Panel B). Figure 4 panels (a) and (b) show that SET has impacts on these outcomes across the distribution, with somewhat larger impacts in the middle of the birth weight distribution and above the median of the gestational age distribution.

4.3 Maternal health outcomes

The maternal health index improves by 0.032 SD for all births, falling just short of significance in the sample of all mothers but it is larger at 0.056 SD and statistically significant for first-time mothers. The SET reform narrows the gap between mothers with IVF and non-IVF births by 36% in this group (observe that the gap is in fact similar for first-time mothers and all mothers).

Components of the index. Estimates for each of the five components are in Table 7. For hypertension and maternal sepsis there is no significant difference in the pre-SET period between IVF and non-IVF users.[22] In line with this, there is no impact of SET. For hemorrhage and hospital re-admission, there is a baseline difference with IVF-mothers having a higher risk by about 4 pp and 1.4 pp respectively, but the SET reform does not lower these risks. However, it has a large impact on the risk of emergency C-section of 2 pp on average and 3 pp for first-time mothers, narrowing the gap between IVF and non-IVF mothers by 42% on average and 60% for first-time mothers. These impacts on emergency C-section remain highly significant after the multiple hypothesis testing correction, irrespective of whether we use the FDR or the Bonferrroni correction.

4.4 Mother's income after birth

An index of maternal income in the three years following birth improves by 0.106 SD on average and by 0.156 for first-time mothers. Consistent with extensive margin fertility having larger im-

[22]The raw difference shown in Table 1 is significant, but now we condition on covariates, in particular mother characteristics.

pacts on income after birth (Lundborg et al., 2014), the IVF/non-IVF gap in income after birth is larger for first-time mothers (-0.163) than for all mothers (-0.125). The estimates suggest that SET narrowed the gap by 85% for the full sample, and nearly closed it for first-time mothers.

Components of the index. Estimates for the three components of the index are in Table 8. After SET, IVF mothers have significantly higher labor earnings and lower usage of benefits and transfers—indeed, SET more than closes the baseline gap. Although we present the coefficients as a percentage of the IVF-non IVF gap in the Tables, using the more familiar benchmark of the baseline sample mean of earnings in the IVF sample, our estimates imply about a 7.2 percent increase. In absolute terms the increase in earnings is around 9600 SEK (1000 USD) per year. Income from parental benefits is not significantly impacted in the full sample, and sickness benefits fall by around 1800 SEK (close to 200 USD). The estimates for labor income and sickness benefits are robust to correcting for multiple hypotheses testing, even when using the conservative Bonferroni correction. For first-time mothers, we similarly estimate a significant increase in labor income (around 8500 SEK, equivalent to 900 USD) and a reduction in sickness benefits of 2100 SEK (around 220 USD). Additionally, in this sample we observe an increase in parental benefits of around 1900 SEK, likely reflecting future childbearing after their first birth. Again, these findings are robust to multiple hypothesis testing corrections.

Intensive vs extensive margin effects on earnings. In Table 9 we observe that the impacts on wage income are virtually all driven by an increase in the intensive margin. When conditioning on observations with non-zero wages only, the impact of SET is estimated to be an increase in average wages of around 9000 SEK or 950 USD (with similar values for first time mothers and higher parity mothers). When examining only the extensive margin (columns 1 and 3), we observe no significant impacts.

Distribution of effects. Figure 5 documents the impact of the SET reform on mother's earnings across the income distribution. Each point and confidence interval corresponds to estimating specification 1, where the outcome is a binary variable indicating that the mother's income exceeds the amount indicated on the horizontal axis. These results suggest that the SET reform had wage impacts across the distribution of wages, with the peak impact close to the mean wage (around 150,000-200,000 SEK), with small effects observable at values as high as 600,000 SEK (around 65,000 USD).

4.5 Robustness Checks

Identifying assumption. Our specification assumes that trends in outcomes among IVF and non-IVF women would have evolved similarly over time in the absence of the reform. We assess this using event study plots, see Figure 6 for the outcome indices. Event studies for each component indicator are presented in Figures A6, A7, and A8. In general, there is no evidence that outcome improvements began before SET. Flexible coefficient models are much more demanding of statistical power than single index models and, for some indicators, the coefficients are noisy but the broad patterns support our design and affirm the main conclusions.[23]

[23] It is also useful to refer back to raw data trends that were shown in Appendix Figures A3, A4 and A5.

We examined sensitivity of the estimates to inclusion of IVF-status-specific trends, see Table A3. We show results with split trends, allowing for a change in slope post-SET for IVF and non-IVF births (Panel A) and with global trends (Panel B). The estimates are very similar in Panels A and B and, in no case, are they significantly different from the estimates without these trends in Table 5. The coefficients are smaller on the indices for child health (0.137 SD relative to 0.189 SD) and mother's income after birth (0.061 SD relative to 0.106 SD) on average. However, conditional on trends, the coefficient on the index for maternal health is larger (0.106 relative to 0.032) and it is now statistically significant not only for first-births but on average. The patterns of results for first-births are broadly similar.

Changes in the composition of IVF mothers. The share of IVF births in all births has increased secularly since the 1990s, tracking changes in technologies, costs, and availability of IVF (Figure 1b). Of potential concern for the interpretation of our estimates is that there was an endogenous change in the composition of IVF-mothers concurrent with the passage of the SET mandate. We examined this in Table A4, where we show post-SET changes in observable characteristics of IVF mothers, namely age, weight, height, education, pre-birth labor income, pre-birth sickness benefits, nationality, smoking, asthma, epilepsy and ulcerative colitis. This is equivalent to a test of balance. We find differences that are small in magnitude but statistically significant for three of eleven characteristics, namely age, education and smoking. Post-SET IVF mothers are slightly older, slightly more educated but also slightly more likely to be smokers.

We address this in a number of ways. First, we consistently show results conditional upon these characteristics. To assess the relevance of unobservable mother characteristics, we postulate that selection based on observables provides information on selection on unobservables, following Altonji et al. (2005). We estimate the magnitude of the omitted variable bias that would have to be present to drive the true impacts of the SET reform to zero.[24] Table A5 presents estimates without maternal controls and the Altonji ratio is in the last row. The large ratios suggest limited selection on observables.

We nevertheless examine estimates conditional on mother fixed effects. Since we present ITT estimates, selection into SET is not a first order concern but this specification also allows for it.[25] We restrict the sample to women with at least two pregnancies, which includes about 50% of all IVF mothers. Since mothers with two pregnancies may not be representative of all IVF mothers, we examine characteristics of IVF mothers with one versus two pregnancies in Table A6. These mothers differ across multiple dimensions, for example, IVF mothers with only one birth have a higher risk of complications than those with two births. For this reason, we estimate the model without mother fixed effects on the reduced sample before we introduce the fixed effects. We can then assess how the coefficients change with the sample independently of how they change with controls for unobserved mother-level heterogeneity. The results are in Table 10. Panel A shows

[24]We compute the ratio of the covariance between unobservables and the SET reform and the covariance between observables and the SET reform using OLS estimates with and without the controls for mother characteristics: $\frac{\alpha_{controls}}{\alpha_{nocontrols}-\alpha_{controls}}$. For a more detailed discussion see Bellows and Miguel (2009).

[25]We showed that the SET reform led to a significant and broadly similar decline in twin births among a broad category of women spanning different education levels, BMI classifications, parity and ages, with the exception of women older than 39. We control flexibly for maternal age. However mother fixed effects will control for the possibility that women select into SET based on unobservables such as multiple previously failed IVF cycles.

results unconditional on mother fixed effects but using the reduced sample and it confirms that the coefficients are similar to those obtained on the full sample. Panel B shows results conditional on mother fixed effects. The coefficients of interest are smaller but result in substantively similar findings.

Third, we re-estimated the model including controls for maternal characteristics in interaction with a linear trend and the indicator for IVF-use.[26] This is a generalization of the specification shown in see Table A3 which includes group (IVF) specific trends. The estimates in Table A7 are broadly similar to those in Table A3. The coefficients of interest are not statistically significantly different, but the coefficients for child and income after birth are slightly smaller and for maternal health are slightly larger, and now statistically significant on average and not only for first births.

Timing and measurement error. Additional robustness checks are presented in Table A8. In Panel A, we remove from the sample births in the two years prior to the SET reform (2001-2002) to allow for increasing openness to or anticipation of SET before it was mandated and/or for measurement error in the date of conception. In Panel B, we remove the region of Skåne because of a regional rule mandating SET as the default starting in 2001 in this region. The results are not sensitive to these changes. In 2005, Sweden started to offer same-sex couples publicly funded access to fertility treatments including IVF. Same-sex couples tend to have higher socioeconomic status (Ahmed et al., 2011a,b) but their children tend to have worse birth outcomes (at least in terms of lower birth weight) (Aldén et al., 2017). Although the number of children born to lesbian parents during 1995-2010 is only 750, we examined the possibility that this legislative change influences our estimates by restricting the sample to conceptions occurring during 1998-2004. The results, in Panel C, are again broadly similar.

The Medical Birth Registry appears to correctly identify 80 to 90 and at least 70 percent of all IVF births, if we compare with reported usage in national IVF data. We may be concerned that some 10 to 30 percent of IVF births are incorrectly reported as non-IVF births, thus contaminating the control group. In practice, given that the size of the "treated" group (IVF users) is much smaller than the size of the "control" group (mothers who do not use IVF), even if the impact of the reform is very large, it is unlikely that the (up to) 30% of mis-classified IVF births will impact averages in the control group in any substantive way. To see this, consider that the number of observed IVF births in the Medical Birth Registry is 21,783, and the number of non-IVF births is 916,110. Inflating the number of IVF births from 70 to 100% implies that 9,356 IVF births are incorrectly classified as non-IVF births. This is only slightly more than 1% of non-IVF births. We provide additional discussion, as well as a calculation of the (small) magnitude of any expected attenuation for the worst case of 30 percent mis-classified in Appendix B.

2SLS estimates. We use the passage of the SET reform to instrument the likelihood of giving birth to a singleton in Table A9. This provides the local average treatment effect on compliers: IVF-using women who had a singleton birth after the policy change, but who would have had twins if SET were not the default policy. Columns 1-2 reproduce the the first stage results, showing F-statistics far exceeding typical weak instrument thresholds. For the full sample, the 2SLS estimates

[26] This allows, for example, that outcomes for more educated women using IVF have a different linear trend compared with outcomes for less educated women using IVF and compared with outcomes for more educated women who do not use IVF.

suggest a strong and significant impact of having a singleton child on child health (1.1 SD of the index) and on maternal income after birth (0.63 SD of the index), while the impact on maternal health (0.186 SD) is large but not significantly different from zero. For first-time mothers, impacts on child health are similar (1.02 SD) and impacts on maternal health (0.31 SD) and income after birth (0.87 SD) are larger.

Heterogeneity in policy impacts. In Appendix Tables A10, A11 and A12 we document reform impacts by maternal composition. These tables examine the impact of the SET reform on the child health index (Appendix Table A10), the maternal income index (Appendix Table A11) and the maternal income index (Appendix Table A12). Heterogeneity is documented by birth order, age group, education level, and BMI classification (following Table 4). In general, for the majority of the groups and the majority of the outcomes, similar reform effects are observed as those documented in the full sample. The exceptions to this are women aged 39 years and over, and women who are underweight according to their BMI, for whom significant impacts are never observed. The lack of impacts in the former group is in line with reform uptake, with exceptions to SET permitted in the case of relatively older women. Underweight women are limited from seeking publicly funded IVF until they meet normal BMI criteria, potentially explaining limited impacts. Both of these effects are in line with the null or limited impacts of the reform to twinning documented in Table 4.

Earnings effects – mechanisms. Maternal wages after birth may improve following SET either directly because the share of twin relative to singleton births declined, or because of the impact that this had in improving child and maternal health. We investigate this here using the conditional decomposition proposed by Gelbach (2016).[27] Results are in Table A13. The total estimated increase in maternal wages (earnings) is 9606 SEK. The decomposition (column 4) shows that changes in child health following SET explain only 300 SEK (or 3%) of the improvement in the wage income, with an even smaller 44 SEK SD (or $<$ 1%) owing to changes in maternal health. This suggests that the direct effect of the drop in twin births (fertility) is the main contributor to improvements in maternal income after SET.

[27]We investigate how much the coefficient associated with the SET reform is diminished when including maternal and child health indicators, $\beta^{uconditional} - \beta^{conditional} = \delta$, in which β^{uc} indicates the unconditional specification excluding child and maternal health and β^{c} expresses the conditional specification including child and maternal health. We can augment this expression by

$$\begin{aligned}\beta_{labor}^{uconditional} - \beta_{labor}^{conditional} \\ &= \Gamma_{labor}^{childhealth}\beta_{labor}^{childhealth} + \Gamma_{labor}^{maternalhealth}\beta_{labor}^{maternalhealth} \\ &= \delta_{labor}^{childhealth} + \delta_{labor}^{maternalhealth} = \delta_{labor} \qquad (3)\end{aligned}$$

where Γ represents each estimate of the SET reform (postSET$\times$IVF) for each potential mechanism as the outcome variable. The coefficient β indicates the estimate of the potential mechanisms as explanatory variables in the full specification with maternal labor outcomes as the dependent variable. The conditional contribution of each component is given by δ, which is computed by multiplying Γ with β.

5 Conclusion

Linking administrative data from several sources at the individual level to create a unique longitudinal data file for all births in Sweden during 1998-2007, we provide a comprehensive examination of causal effects of a 2003 reform that mandated single embryo transfer (SET) in IVF fertility treatment, displacing the default of double embryo transfers (DET). We demonstrate a sharp drop in the share of twin births to IVF mothers of all education, height, weight and age categories, with the exception of women over the age of 39, possibly because they were exempt and used DET. In line with this we find across the board improvements in child and maternal health and in maternal income in the three years following birth. These improvements are large.

Improved child health and lower rates of emergency C-Section have a direct impact on the financial costs of IVF. In addition, we expect that improved child health along numerous dimensions will generate higher cognitive attainment, employment, income and life expectancy for IVF births, and that women who return to work will permanently be on a higher earnings trajectory. Thus the the benefits of the SET mandate are likely to be larger than we document. This is important as IVF is now a key feature of the reproductive landscape and likely to continue to increase, especially as it becomes more readily accessible to women in poorer countries.

References

Abramowitz, J. (2014): "Turning back the ticking clock: the effect of increased affordability of assisted reproductive technology on women's marriage timing," *Journal of Population Economics*, 27, 603–633.

——— (2017): "Assisted Reproductive Technology and Women's Timing of Marriage and Childbearing," *Journal of Family and Economic Issues*, 38, 100–117.

Adda, J., C. Dustmann, and K. Stevens (2017): "The career costs of children," *Journal of Political Economy*, 125, 293–337.

Ahmed, A. M., L. Andersson, and M. Hammarstedt (2011a): "Inter-and Intra-Household Earnings Differentials among Homosexual and Heterosexual Couples," *British Journal of Industrial Relations*, 49.

Ahmed, A. M., L. Andersson, M. Hammarstedt, et al. (2011b): "Sexual orientation and occupational rank," *Economics Bulletin*, 31, 2422–2433.

Aldén, L., A. Bjorklund, and M. Hammarstedt (2017): "Early Health and School Outcomes for Children with Lesbian Parents: Evidence from Sweden," .

Alderman, H., M. Lokshin, and S. Radyakin (2011): "Tall claims: Mortality selection and the height of children in India," *Economics & Human Biology*, 9, 393–406.

Alm, A. (2010): "Olika regler i olika landsting," *Gefle Dagblad*.

Almond, D., J. J. Doyle, A. E. Kowalski, and H. Williams (2010): "Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns," *The Quarterly Journal of Economics*, 125, 591–634.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005): "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of political economy*, 113, 151–184.

Anderson, M. L. (2008): "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American statistical Association*, 103, 1481–1495.

Anell, A. (2008): "The Swedish health care system," *New York: The Commonwealth Fund*.

Bailey, M. J. and J. M. Lindo (2017): "Access and Use of Contraception and Its Effects on Women's Outcomes in the US," Tech. rep., National Bureau of Economic Research.

Bamfo, J. E. (2013): "Managing the risks of sepsis in pregnancy," *Best practice & research Clinical obstetrics & gynaecology*, 27, 583–595.

Behrman, J. R. and M. R. Rosenzweig (2004): "Returns to birthweight," *Review of Economics and statistics*, 86, 586–601.

Bellows, J. and E. Miguel (2009): "War and local collective action in Sierra Leone," *Journal of Public Economics*, 93, 1144–1157.

Benjamini, Y. and Y. Hochberg (1995): "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the royal statistical society. Series B (Methodological)*, 289–300.

Bergh, T., A. Ericson, T. Hillensjö, K. Nygren, and U.-B. Wennerholm (1999): "Deliveries and children born after in-vitro fertilisation in Sweden 1982–95: a retrospective cohort study," *The Lancet*, 354, 1579–1585.

Bhalotra, S. and D. Clarke (2019a): "The Twin Instrument: Fertility and Human Capital Investment," *Journal of the European Economic Association*, jvz058.

——— (2019b): "Twin Birth and Maternal Condition," *Review of Economics and Statistics*, 101, 853–864.

Bhalotra, S., M. Karlsson, and T. Nilsson (2017): "Infant health and longevity: Evidence from a historical intervention in Sweden," *Journal of the European Economic Association*, jvx028.

Bhalotra, S. R., A. Venkataramani, and S. Walther (2018): "Fertility and Labor Market Responses to Reductions in Mortality," IZA Discussion Papers 11716, Institute of Labor Economics (IZA).

Bharadwaj, P., K. V. Løken, and C. Neilson (2013): "Early Life Health Interventions and Academic Achievement," *American Economic Review*, 103, 1862–91.

Bitler, M. P. (2008): "Effects of increased access to infertility treatment on infant and child health: Evidence from health insurance mandates," *Unpublished manuscript*.

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

——— (2012): "Utilization of Infertility Treatments: The Effects of Insurance Mandates," *Demography*, 49, 125–149.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2007): "From the cradle to the labor market? The effect of birth weight on adult outcomes," *The Quarterly Journal of Economics*, 122, 409–439.

Bloom, D. E., M. Kuhn, and K. Prettner (2015): "The Contribution of Female Health to Economic Development," Working Paper 21411, National Bureau of Economic Research.

Bonferroni, C. E. (1935): "Il calcolo delle assicurazioni su gruppi di teste," *Studi in onore del professore salvatore ortu carboni*, 13–60.

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

Bundorf, M. K., M. Henne, and L. Baker (2007): "Mandated Health Insurance Benefits and the Utilization and Outcomes of Infertility Treatments," Working Paper 12820, National Bureau of Economic Research.

Criniti, A., A. Thyer, G. Chow, P. Lin, N. Klein, and M. Soules (2005): "Elective single blastocyst transfer reduces twin rates without compromising pregnancy rates," *Fertility and sterility*, 84, 1613–1619.

Currie, J. and H. Schwandt (2013): "Within-mother analysis of seasonal patterns in health at birth," *Proceedings of the National Academy of Sciences*, 110, 12265–12270.

de Mouzon, J., V. Goossens, S. Bhattacharya, J. Castilla, A. Ferraretti, V. Korsak, M. Kupka, K.-G. Nygren, and A. Nyboe Andersen (2010): "Assisted reproductive technology in Europe, 2006: results generated from European registers by ESHRE," *Human Reproduction*, 25, 1851–1862.

Gelbach, J. B. (2016): "When Do Covariates Matter? And Which Ones, and How Much?" *Journal of Labor Economics*, 34, 509–543.

Gelbaya, T. A., I. Tsoumpou, and L. G. Nardo (2010): "The likelihood of live birth and multiple birth after single versus double embryo transfer at the cleavage stage: a systematic review and meta-analysis," *Fertility and sterility*, 94, 936–945.

Gerris, J., E. Van Royen, D. De Neubourg, K. Mangelschots, M. Valkenburg, and G. Ryckaert (2001): "Impact of single embryo transfer on the overall and twin-pregnancy rates of an IVF/ICSI programme," *Reproductive BioMedicine Online*, 2, 172–177.

Hall, J. G. (2003): "Twinning," *The Lancet*, 362, 735–743.

Hamilton, B. H., E. Jungheim, B. McManus, and J. Pantano (2018): "Health Care Access, Costs, and Treatment Dynamics: Evidence from In Vitro Fertilization," *American Economic Review*, 108, 3725–77.

Horowitz, J. L. and C. F. Manski (1995): "Identification and robustness with contaminated and corrupted data," *Econometrica: Journal of the Econometric Society*, 281–302.

Kalra, S. K. and K. T. Barnhart (2011): "In vitro fertilization and adverse childhood outcomes: what we know, where we are going, and how we will get there. A glimpse into what lies behind and beckons ahead," *Fertility and sterility*, 95, 1887–1889.

Karlström, P. O. and C. Bergh (2007): "Reducing the number of embryos transferred in Sweden-impact on delivery and multiple birth rates," *Human Reproduction*, 22, 2202–2207.

Kleven, H., C. Landais, and J. E. Søgaard (2019): "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, 11, 181–209.

Kroeger, S. and G. La Mattina (2017): "Assisted reproductive technology and women's choice to pursue professional careers," *Journal of Population Economics*, 30, 723–769.

Kutlu, P., O. Atvar, O. Vanlioglu, U. Kutlu, A. Arici, S. Yilmaz, E. Yilmaz, N. Delikara, F. Bener, A. Kamar, O. Alpak, and U. Ozekici (2011): "Effect of the new legislation and single-embryo transfer policy in Turkey on assisted reproduction outcomes: preliminary results," *Reproductive BioMedicine Online*, 22, 208 – 214.

Lee, D. S. (2009): "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," *The Review of Economic Studies*, 76, 1071–1102.

Lemos, E. V., D. Zhang, B. J. V. Voorhis, and X. H. Hu (2013): "Healthcare expenses associated with multiple vs singleton pregnancies in the United States," *American Journal of Obstetrics & Gynecology*, 209, 586.e1–586.e11.

Lukassen, H. M., D. Braat, A. M. Wetzels, G. A. Zielhuis, E. M. Adang, E. Scheenjes, and J. A. Kremer (2005): "Two cycles with single embryo transfer versus one cycle with double embryo transfer: a randomized controlled trial," *Human Reproduction*, 20, 702–708.

Lundborg, P., E. Plug, and A. W. Rasmussen (2014): "Fertility Effects on Female Labor Supply: IV Evidence from IVF Treatments," .

Lundin, K. and C. Bergh (2007): "Cumulative impact of adding frozen–thawed cycles to single versus double fresh embryo transfers," *Reproductive biomedicine online*, 15, 76–82.

Machado, M. P. and A. Sanz-de Galdeano (2015): "Coverage of infertility treatment and fertility outcomes," *SERIEs*, 6, 407–439.

Maheshwari, A., S. Griffiths, and S. Bhattacharya (2011): "Global variations in the uptake of single embryo transfer," *Human Reproduction Update*, 17, 107–120.

Oreopoulos, P., M. Stabile, R. Walld, and L. L. Roos (2008): "Short-, medium-, and long-term consequences of poor infant health an analysis using siblings and twins," *Journal of human Resources*, 43, 88–138.

Pinborg, A., U.-B. Wennerholm, L. Romundstad, A. Loft, K. Aittomaki, V. Söderström-Anttila, K. Nygren, J. Hazekamp, and C. Bergh (2013): "Why do singletons conceived after assisted reproduction technology have adverse perinatal outcome? Systematic review and meta-analysis," *Human reproduction update*, 19, 87–104.

Pinckney-Clark, E., F. I. Sharara, J. M. Franasiak, P. W. Heiser, and T. Tobias (2016): "Promoting the use of elective single embryo transfer in clinical practice," *Fertility Research and Practice*, 2, 1.

Practice Committee of the Society for Assisted Reproductive Technology and Practice Committee of the American Society for Reproductive Medicine and others (2012): "Elective single-embryo transfer," *Fertility and Sterility*, 97, 835–842.

Rainer, H., G. Selvaretnam, and D. Ulph (2011): “Assisted reproductive technologies (ART) in a model of fertility choice,” *Journal of Population Economics*, 24, 1101–1132.

Saldeen, P. and P. Sundström (2005): “Would legislation imposing single embryo transfer be a feasible way to reduce the rate of multiple pregnancies after IVF treatment?” *Human Reproduction*, 20, 4–8.

Sazonova, A., K. Källen, A. Thurin-Kjellberg, U.-B. Wennerholm, and C. Bergh (2011): “Obstetric outcome after in vitro fertilization with single or double embryo transfer,” *Human reproduction*, 26, 442–450.

——— (2013): “Neonatal and maternal outcomes comparing women undergoing two in vitro fertilization (IVF) singleton pregnancies and women undergoing one IVF twin pregnancy.” *Fertility and Sterility*, 99, 731–737.

SKL (2016): “Rekommendation om enhetlighet i landstingens och regionernas erbjudande av offentlig finansierad assisterad befruktning,” *Sveriges Kommuner och Landsting*.

Socialstyrelsen (2006): “Assisterad befruktning 2003, Assisted reproduction, Results of treatment 2003,” *Socialstyrelsen*.

Socialstyrelsen (2013): “Graviditeter, förlossningar och nyfödda barn. Medicinska födelseregistret 1973–2011. Assisterad befruktning 1991–2010,” *Socialstyrelsen*.

Socialstyrelsen (2016): “Assisterad befruktning med donerade könsceller Nationellt kunskapsstöd,” *Socialstyrelsen*.

Thurin, A., J. Hausken, T. Hillensjö, B. Jablonowska, A. Pinborg, A. Strandell, and C. Bergh (2004): “Elective single-embryo transfer versus double-embryo transfer in in vitro fertilization,” *New England Journal of Medicine*, 351, 2392–2402.

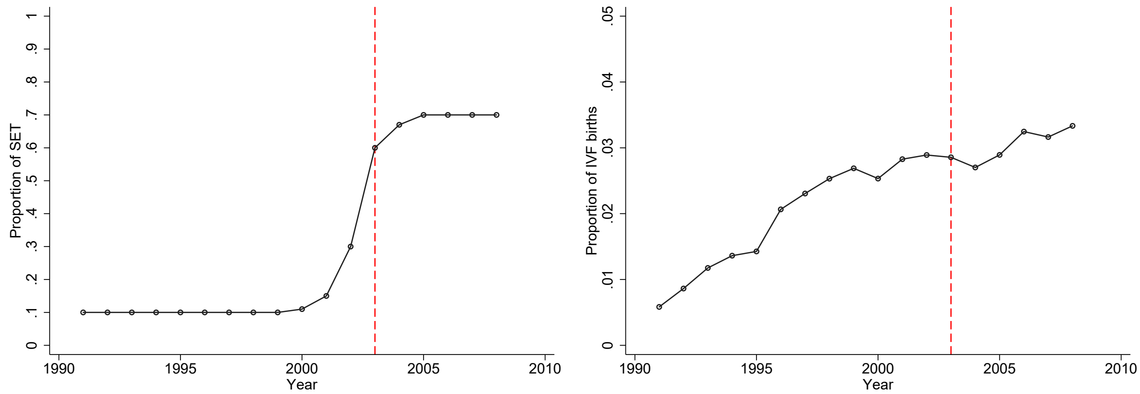
Vilska, S., A. Tiitinen, C. Hyden-Granskog, and O. Hovatta (1999): “Elective transfer of one embryo results in an acceptable pregnancy rate and eliminates the risk of multiple birth,” *Human Reproduction*, 14, 2392–2395.

Zegers-Hochschild, F., G. D. Adamson, J. de Mouzon, O. Ishihara, R. Mansour, K. Nygren, E. Sullivan, and S. Van der Poel (2009): “The international committee for monitoring assisted reproductive technology (ICMART) and the world health organization (WHO) revised glossary on ART terminology, 2009,” *Human Reproduction*, dep343.

Figures and Tables

Figure 1: Trends in SET and proportion of IVF births

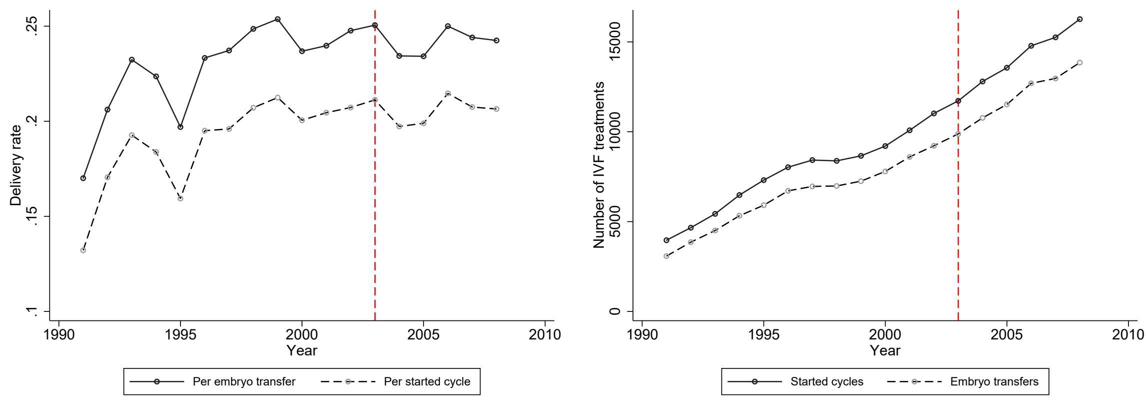
(a) Single embryo transfer **(b)** Proportion of IVF births



Notes: Annual trends in SET and the proportion of IVF births are based on aggregate data collected from annual reports by the Swedish National Board of Health and Welfare and presented in Figures 1a and 1b. The red vertical line indicates the year of the SET reform.

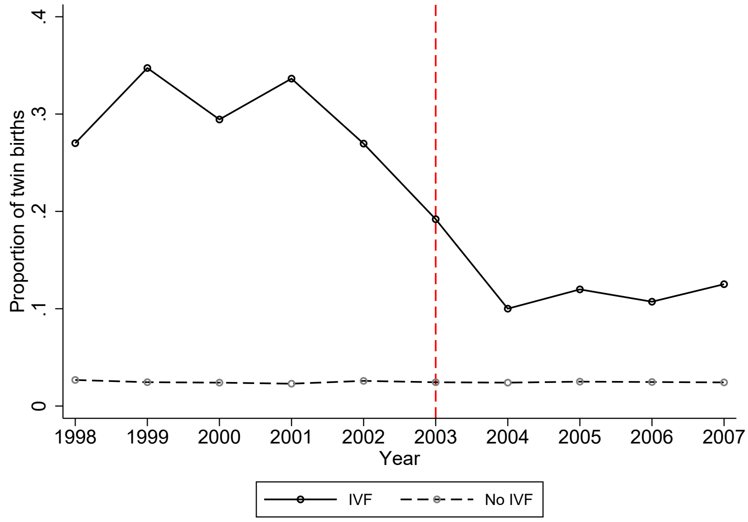
Figure 2: Trends in delivery rate and IVF treatments

(a) Deliveries per embryo transfer and cycle **(b)** Number of IVF treatments



Notes: Annual trends in deliveries per transfer/cycle and the number of IVF treatments are based on aggregate data collected from annual reports by the Swedish National Board of Health and Welfare and presented in Figures 2a and 2b. The red vertical line indicates the year of the SET reform.

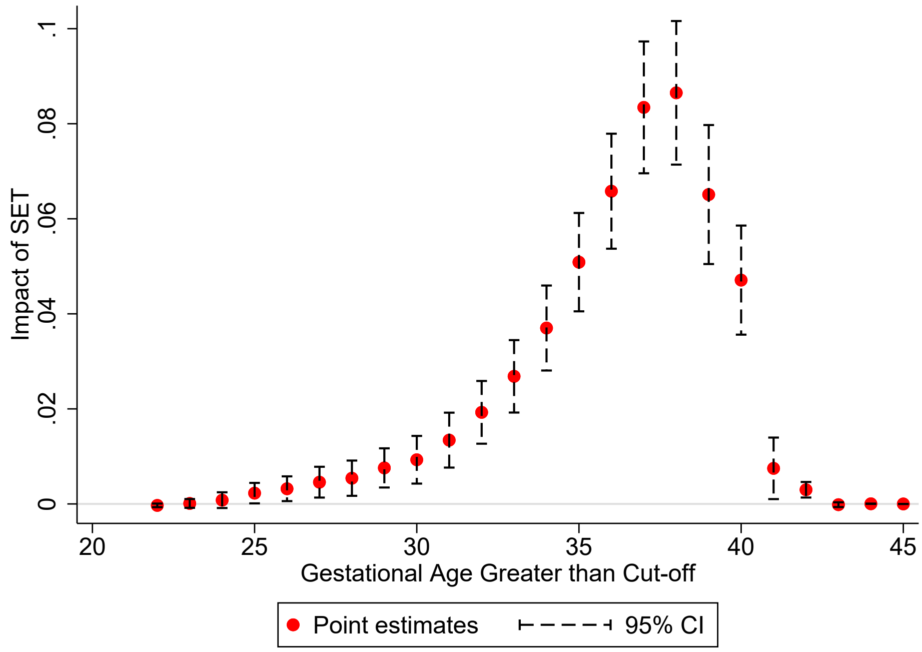
Figure 3: Trends in twin rates



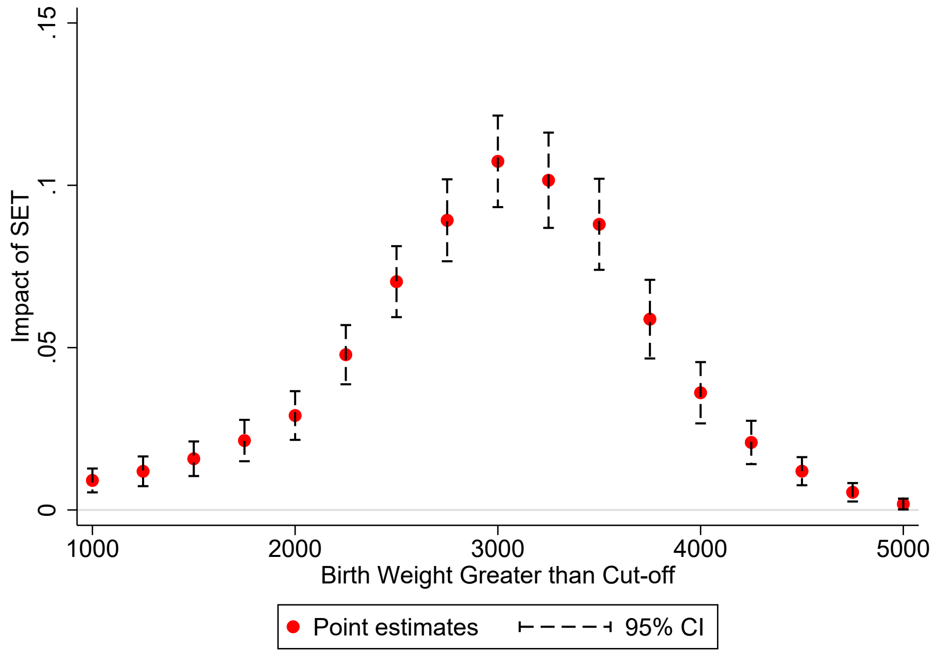
Notes: Annual trends in twin births are presented for conceptions with and without IVF treatment. Data are obtained from the Swedish Medical Birth Registry (microdata records). The red vertical line indicates the year of the SET reform.

Figure 4: Distributional Impacts of the SET Reform on Health at Birth

(a) Gestational age

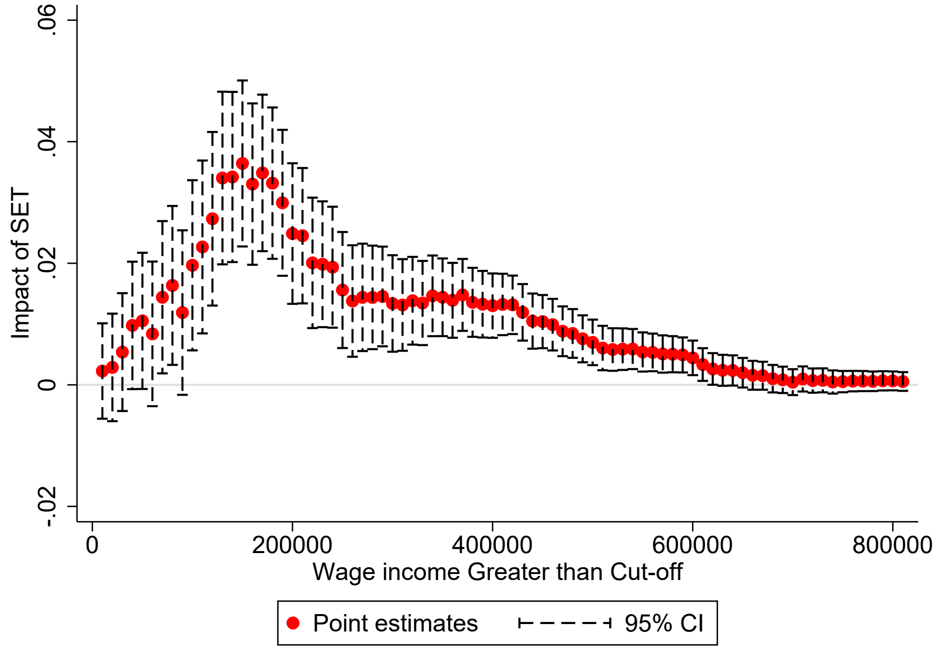


(b) Birthweight



Notes: Each point estimate and 95% confidence interval (CI) refers to the coefficient and CI on $PostSET \times IVF$ from equation 1, where in each case the dependent variable of interest is a binary variable, taking the value of 1 if the birth exceeds the gestational age (panel (a)), or birthweight (panel (b)) indicated on the horizontal axis, and 0 otherwise.

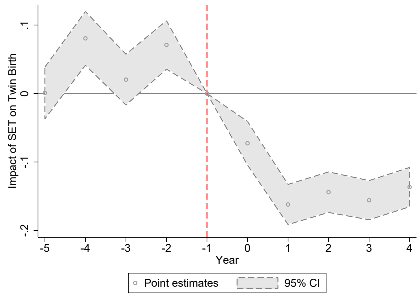
Figure 5: Distributional Impacts of the SET Reform on Wage Income After Birth



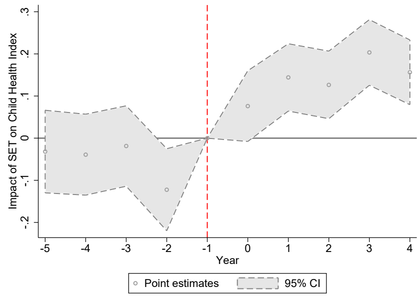
Notes: Each point estimate and 95% confidence interval (CI) refers to the coefficient and CI on $PostSET \times IVF$ from equation 1, where in each case the dependent variable of interest is a binary variable, taking the value of 1 if the mother's average wage income in the three years following birth exceeds the amount (in SEK) indicated on the horizontal axis, and 0 otherwise.

Figure 6: Event studies: main results

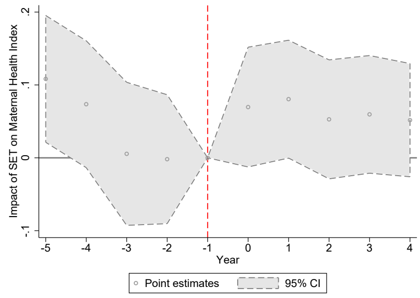
(a) Twin birth, full sample



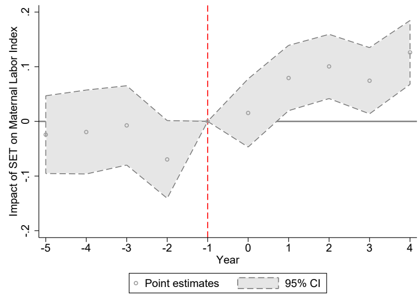
(b) Child health index, full sample



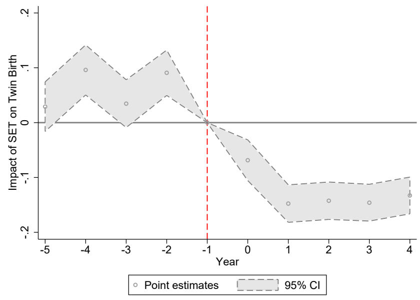
(c) Maternal health index, full sample



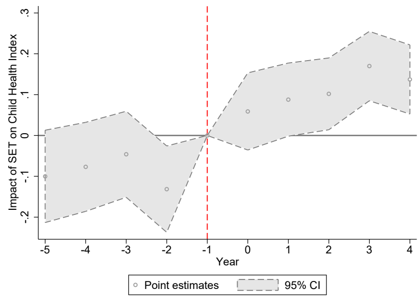
(d) Labor market index, full sample



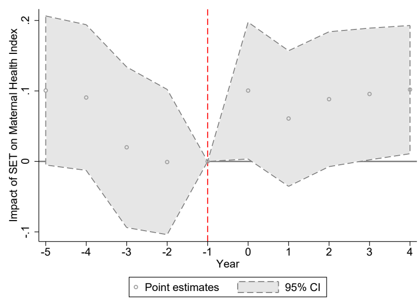
(e) Twin birth, 1st-time mothers



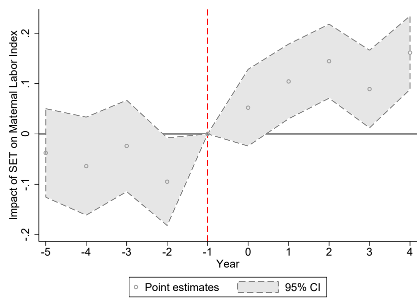
(f) Child health index, 1st-time mothers



(g) Maternal health index, 1st-time-time mothers



(h) Labor market index, 1st-time-time mothers



Notes: Event studies are based on microdata obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007. Each figure presents coefficients of the full set of reform lags and leads interacted with IVF births, as per specification 2. The red-vertical line represents the year of the SET reform, and year -1 is the omitted reference period. A full set of maternal controls and fixed effects are included in all regressions (as described in Table 3). Standard errors are clustered by mother.

Table 1: Summary statistics

	IVF Mean (1)	*Non-IVF* Mean (2)	P-values (3)
Panel A: Maternal characteristics			
Twin birth	0.303	0.025	0.000
Age	33.058	29.704	0.000
Height	167.339	166.378	0.000
Weight	68.681	67.607	0.000
BMI	24.514	24.421	0.055
Tobacco use before pregnancy	0.058	0.115	0.000
Tobacco use during pregnancy	0.028	0.076	0.000
Asthma	0.062	0.063	0.581
Ulcerative colitis	0.010	0.005	0.000
Epilepsy	0.003	0.003	0.234
Panel B: Maternal outcomes			
Emergency C-section	0.173	0.082	0.000
Maternal sepsis	0.003	0.002	0.001
Postpartum hemorrhage	0.111	0.058	0.000
Post-birth hospitalization	0.065	0.053	0.000
Hypertension	0.003	0.002	0.205
Maternal health index	-0.257	0.014	0.000
Education	4.530	4.357	0.000
Labor income	134267	110175	0.000
Sickness benefits	13891	9887	0.000
Parental benefits	56506	49943	0.000
Maternal labor index	-0.057	-0.150	0.000
Panel C: Child outcomes			
Apgar score	9.585	9.727	0.000
Apgar score < 7	0.021	0.011	0.000
Birth weight	3197.993	3547.714	0.000
Gestational age (weeks)	38.309	39.339	0.000
Head circumference	34.348	34.906	0.000
Length (centimeters)	49.269	50.420	0.000
Gender (male)	0.512	0.515	0.645
Breech presentation	0.115	0.037	0.000
Malformation	0.045	0.035	0.000
Infant mortality $\times$ 1000	6.640	2.970	0.000
Under 5 mortality $\times$ 1000	8.103	3.806	0.000
Hospitalization ages 0-1	0.255	0.167	0.000
Hospitalization ages 1-4	0.155	0.137	0.000
Child health index	-0.393	-0.003	0.000

Notes to Table 1. Summary statistics are presented for data obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies. The sample includes 8,886 IVF and 428,294 non-IVF births for the time period 1998-2002 (the period preceding the implementation of the SET reform). There is minor variation in the total number of observations between variables given a small number of missing outcomes for certain variables. Mean values are presented, and p-values in column 3 refer to two-tailed t-tests of equality of means between IVF and non-IVF births.

Table 2: Summary statistics, twins and singletons

	Twins			*Singletons*		
	IVF	Non-IVF	Difference	IVF	Non-IVF	Difference
	Mean	Mean	P-values	Mean	Mean	P-values
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Maternal characteristics						
Age	32.659	30.721	0.000	33.231	29.678	0.000
Height	167.632	167.156	0.001	167.212	166.359	0.000
Weight	68.829	69.060	0.428	68.617	67.571	0.000
BMI	24.449	24.710	0.008	24.542	24.414	0.027
Tobacco use before pregnancy	0.044	0.128	0.000	0.064	0.114	0.000
Tobacco use during pregnancy	0.017	0.060	0.000	0.033	0.076	0.000
Asthma	0.065	0.058	0.215	0.060	0.063	0.348
Ulcerative colitis	0.010	0.005	0.000	0.010	0.005	0.000
Epilepsy	0.002	0.003	0.775	0.003	0.003	0.452
Panel B: Maternal outcomes						
Emergency C-section	0.275	0.249	0.007	0.128	0.077	0.000
Maternal sepsis	0.005	0.006	0.702	0.003	0.002	0.070
Postpartum hemorrhage	0.168	0.143	0.001	0.086	0.055	0.000
Post-birth hospitalization	0.077	0.068	0.111	0.060	0.052	0.009
Hypertension	0.003	0.003	0.470	0.003	0.002	0.352
Maternal health index	-0.547	-0.452	0.002	-0.130	0.026	0.000
Education	4.542	4.367	0.000	4.524	4.357	0.000
Labor income	131,756	109,289	0.000	135,358	110,198	0.000
Sickness benefits	17,827	15,336	0.000	12,182	9,749	0.000
Parental benefits	61,218	55,815	0.000	54,459	49,793	0.000
Maternal labor index	-0.106	-0.221	0.000	-0.036	-0.149	0.000
Panel C: Child outcomes						
Apgar score	9.429	9.476	0.072	9.651	9.733	0.000
Apgar score < 7	0.033	0.033	0.909	0.015	0.011	0.001
Birth weight	2,563	2,597	0.014	3,470	3,571	0.000
Gestational age (weeks)	36.171	36.239	0.301	39.237	39.418	0.000
Head circumference	33.158	33.146	0.808	34.814	34.947	0.000
Length (centimeters)	46.716	46.776	0.440	50.301	50.506	0.000
Gender (male)	0.499	0.505	0.602	0.518	0.515	0.661
Breech presentation	0.266	0.268	0.788	0.050	0.031	0.000
Malformation	0.049	0.042	0.083	0.043	0.034	0.000
Infant mortality $\times$ 1000	13.016	14.852	0.476	3.873	2.667	0.069
Under 5 mortality $\times$ 1000	14.132	16.450	0.392	5.487	3.484	0.008
Hospitalization ages 0-1	0.395	0.369	0.012	0.194	0.162	0.000
Hospitalization ages 1-4	0.169	0.157	0.138	0.150	0.137	0.003
Child index	-1.044	-1.006	0.298	-0.111	0.023	0.000

Notes to Table 2. Summary statistics are presented separately for twin and non-twin births, split by each birth's IVF status. Data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies. The sample includes IVF and non-IVF births, for the pre-treatment period of 1998-2002. Mean values are presented, as p-values corresponding to two-tailed t-tests of equality of values between IVF and non-IVF users.

Table 3: Probability of twinning

	Full sample		*First-time mothers*	
	(1) *Twin birth*	(2) *Twin birth*	(3) *Twin birth*	(4) *Twin birth*
postSET×IVF	-0.173***	-0.168***	-0.177***	-0.177***
	(0.007)	(0.007)	(0.009)	(0.009)
IVF	0.276***	0.268***	0.274***	0.268***
	(0.006)	(0.006)	(0.007)	(0.007)
Effect size	-62 %	-63 %	-65 %	-66 %
Fixed effects	NO	YES	NO	YES
Controls	NO	YES	NO	YES
R-Squared	0.038	0.062	0.064	0.067
Observations	937893	937893	414182	414182
Mean of dep. var.	0.029	0.029	0.029	0.029
Control mean	0.027	0.027	0.027	0.027
Control sd	0.163	0.163	0.162	0.162

Notes to Table 3. Each column presents a separate OLS regression with DiD estimates of the impact of the SET reform on the probability of twin birth for the full sample (columns 1-2) and first-time mothers (columns 3-4). In columns 1 and 3, controls are excluded. In columns 2 and 4, a full set of maternal and child controls and fixed effects are included (age, birth order, country fixed effects along with maternal characteristics such as weight, height, smoking, and binary variables for missing values). All data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007. Estimated date of conception fixed effects are included in all regressions. Standard errors are clustered by mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 4: Probability of twinning, heterogeneous effects

Dependent variable: twin birth

Panel A Sub-Sample:	Birth order			Age groups				
	(1) Birth order 1	(2) Birth order 2	(3) Birth order ≥ 3	(4) Ages < 25	(5) Ages 25-29	(6) Ages 30-34	(7) Ages 35-38	(8) Ages ≥ 39
postSET×IVF	-0.178***	-0.160***	-0.0950***	-0.108**	-0.223***	-0.200***	-0.128***	-0.0235
	(0.0087)	(0.0153)	(0.0329)	(0.0456)	(0.0182)	(0.0113)	(0.0125)	(0.0231)
IVF	0.268***	0.276***	0.212***	0.174***	0.308***	0.288***	0.235***	0.140***
	(0.0074)	(0.0129)	(0.0264)	(0.0361)	(0.0152)	(0.0096)	(0.0105)	(0.0192)
Effect size	-66 %	-58 %	-45 %	-62 %	-72 %	-69 %	-55 %	-17 %
Observations	414182	341724	181347	132592	290706	331919	153399	29275
R^2	0.066	0.045	0.047	0.085	0.067	0.079	0.104	0.254
Mean of dep. var.	0.0290	0.0278	0.0264	0.0162	0.0242	0.0316	0.0401	0.0373

Panel B Sub-Sample:	Education			BMI classification			
	(1) Primary education	(2) Secondary education	(3) Tertiary education	(4) Under weight BMI < 18.5	(5) Normal weight BMI 18.5-24	(6) Overweight BMI 25-29	(7) Obesity BMI ≥ 30
postSET×IVF	-0.146***	-0.162***	-0.183***	-0.0940**	-0.179***	-0.151***	-0.151***
	(0.0154)	(0.0116)	(0.0123)	(0.0479)	(0.0098)	(0.0148)	(0.0224)
IVF	0.266***	0.257***	0.275***	0.207***	0.270***	0.270***	0.236***
	(0.0118)	(0.0098)	(0.0108)	(0.0382)	(0.0084)	(0.0123)	(0.0195)
Effect size	-55 %	-63 %	-67 %	-45 %	-66 %	-56 %	-64 %
Observations	236541	383284	312432	19433	498958	211379	90396
R^2	0.089	0.067	0.078	0.344	0.069	0.088	0.123
Mean of dep. var.	0.0291	0.0280	0.0294	0.0226	0.0272	0.0301	0.0300

Notes to Table 4. Refer to notes to Table 3. Identical regressions are estimated, however now separating for different sub-samples of birth order, age groups, education attainments and BMI classifications. All regressions including county, estimated date of conception, birth order and age fixed effects and binary variables for missing values. Standard errors are clustered by mother. * p<0.1, ** p<0.05, *** p<0.01.

Table 5: Effects of SET on child and maternal outcomes

	Full sample			*First-time mothers*		
	(1) Child health index	(2) Maternal health index	(3) Maternal labor index	(4) Child health index	(5) Maternal health index	(6) Maternal labor index
postSET×IVF	0.189***	0.032	0.106***	0.184***	0.056**	0.156***
	(0.019)	(0.020)	(0.016)	(0.022)	(0.023)	(0.019)
IVF	-0.355***	-0.165***	-0.125***	-0.319***	-0.156***	-0.163***
	(0.016)	(0.015)	(0.013)	(0.018)	(0.018)	(0.016)
Effect size	-53 %	-19 %	-85 %	-58 %	-36 %	-96 %
R-Squared	0.019	0.023	0.196	0.021	0.027	0.198
Observations	937893	937893	936777	414180	414180	413652
Mean of dep. var.	-0.003	-0.003	0.005	-0.003	-0.004	0.006
Control mean	0.000	-0.000	-0.000	0.000	0.000	-0.000
Control sd	1.000	1.000	1.000	1.000	1.000	1.000

Note to Table 5. Each column presents a separate OLS regression with DiD estimates of the impact of the SET reform on a child health index (columns 1 and 4), maternal health index (columns 2 and 5), or maternal labor market index (columns 3 and 6). Panel A presents estimates for the full sample and Panel B for a sub-sample of first-time mothers. A full set of maternal controls and fixed effects are included in all regressions (as described in Table 3). Standard errors are clustered by mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 6: Effects of SET on child health- by component

Panel A: *Full sample*

	(1) Apgar score	(2) Apgar score <7	(3) Birth weight	(4) Birth length	(5) Head circumference	(6) Gestation	(7) Male	(8) Infant mortality rate	(9) Under five mortality	(10) Breech position	(11) Child hospitalization ages 0-1	(12) Child hospitalization ages 1-4	(13) Fetal malformation
postSET×IVF	0.061***	-0.005**	175.119***	0.605***	0.235***	0.539***	0.005	-2.993***	-3.492***	-0.036***	-0.045***	-0.002	-0.001
	(0.014)	(0.002)	(11.343)	(0.052)	(0.032)	(0.045)	(0.007)	(1.062)	(1.163)	(0.004)	(0.006)	(0.005)	(0.003)
IVF	-0.082***	0.006***	-307.348***	-1.121***	-0.547***	-1.037***	-0.003	3.748***	4.457***	0.063***	0.088***	0.024***	0.007***
	(0.011)	(0.002)	(9.283)	(0.042)	(0.026)	(0.037)	(0.005)	(0.932)	(1.017)	(0.004)	(0.005)	(0.004)	(0.002)
Effect size	-74 %	-83 %	-57 %	-54 %	-43 %	-52 %	-167 %	-80 %	-78 %	-57 %	-51 %	-8 %	-14 %
FDR p-value (Treat)	0.000	0.067	0.000	0.000	0.000	0.000	0.304	0.013	0.012	0.000	0.000	0.753	0.517
Bonferroni p-value (Treat)	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.211	0.182	0.000	0.000	1.000	1.000
R-Squared	0.021	0.008	0.093	0.075	0.057	0.024	0.004	0.006	0.005	0.012	0.019	0.009	0.005
Observations	930302	930302	935714	925477	894087	937893	937870	937893	937893	937893	937893	937893	937893
Mean of dep. var.	9.729	0.012	3528.062	50.360	34.918	39.297	0.514	2.755	3.483	0.039	0.165	0.132	0.036

Panel B: *First-time mothers*

	(1) Apgar score	(2) Apgar score <7	(3) Birth weight	(4) Birth length	(5) Head circumference	(6) Gestation	(7) Male	(8) Infant mortality rate	(9) Under five mortality	(10) Breech position	(11) Child hospitalization ages 0-1	(12) Child hospitalization ages 1-4	(13) Fetal malformation
postSET×IVF	0.074***	-0.005**	185.631***	0.646***	0.242***	0.591***	0.010	-3.127***	-3.541***	-0.031***	-0.049***	0.002	-0.003
	(0.017)	(0.002)	(13.625)	(0.063)	(0.039)	(0.055)	(0.009)	(1.181)	(1.313)	(0.005)	(0.008)	(0.006)	(0.004)
IVF	-0.074***	0.004**	-293.378***	-1.120***	-0.525***	-1.074***	-0.008	3.261***	3.989***	0.057***	0.087***	0.018***	0.008***
	(0.014)	(0.002)	(11.176)	(0.052)	(0.032)	(0.046)	(0.007)	(1.050)	(1.155)	(0.004)	(0.006)	(0.005)	(0.003)
Effect size	-100 %	-125 %	-63 %	-58 %	-46 %	-55 %	-125 %	-96 %	-89 %	-54 %	-56 %	11 %	-38 %
FDR p-value (Treat)	0.000	0.067	0.000	0.000	0.000	0.000	0.304	0.013	0.012	0.000	0.000	0.753	0.517
Bonferroni p-value (Treat)	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.211	0.182	0.000	0.000	1.000	1.000
R-Squared	0.023	0.012	0.066	0.063	0.045	0.027	0.009	0.010	0.010	0.014	0.029	0.014	0.010
Observations	411406	411406	413106	407269	393470	414180	414167	414180	414180	414180	414180	414180	414180
Mean of dep. var.	9.671	0.015	3431.773	50.133	34.739	39.299	0.515	2.764	3.530	0.052	0.168	0.132	0.038

Notes to Table 6. Refer to notes to Table 5. This table presents identical regression results, however now for each component of the child health index. Given the multiplicity of outcome variables considered, both FDR and Bonferroni corrected p-values are reported in table footers. These multiple hypothesis corrections are made considering multiple outcomes in child health, maternal health and maternal income indexes jointly. * p<0.1, ** p<0.05, *** p<0.01.

Table 7: Effects of SET on maternal health- by component

	Panel A: *Full sample*				
	(1) Emergency C-section	(2) Hemorrhage	(3) Maternal sepsis	(4) Hospital re-admission	(5) Hypertension
postSET×IVF	-0.020***	-0.001	0.000	-0.004	0.001
	(0.006)	(0.005)	(0.001)	(0.004)	(0.001)
IVF	0.048***	0.039***	0.001	0.014***	-0.001
	(0.005)	(0.004)	(0.001)	(0.003)	(0.001)
Effect size	-42 %	-3 %	0 %	-29 %	-100 %
FDR p-value (Treat)	0.000	0.282	0.645	0.362	0.532
Bonferroni p-value (Treat)	0.003	1.000	1.000	1.000	1.000
R-Squared	0.040	0.011	0.006	0.009	0.007
Observations	937884	937893	937893	888342	937893
Mean of dep. var.	0.085	0.063	0.002	0.051	0.004

	Panel B: *First-time mothers*				
	(1) Emergency C-section	(2) Hemorrhage	(3) Maternal sepsis	(4) Hospital re-admission	(5) Hypertension
postSET×IVF	-0.029***	-0.008	-0.001	-0.005	0.001
	(0.007)	(0.006)	(0.001)	(0.005)	(0.001)
IVF	0.048***	0.041***	0.001	0.014***	-0.001
	(0.006)	(0.005)	(0.001)	(0.004)	(0.001)
Effect size	-60 %	-20 %	-100 %	-36 %	-100 %
FDR p-value (Treat)	0.000	0.282	0.645	0.362	0.532
Bonferroni p-value (Treat)	0.003	1.000	1.000	1.000	1.000
R-Squared	0.041	0.017	0.011	0.014	0.012
Observations	414177	414180	414180	409619	414180
Mean of dep. var.	0.123	0.074	0.002	0.053	0.003

Notes to Table 7. Refer to notes to Table 5. This table presents identical regression results, however now for each component of the maternal health index. Given the multiplicity of outcome variables considered, both FDR and Bonferroni corrected p-values are reported in table footers. These multiple hypothesis corrections are made considering multiple outcomes in child health, maternal health and maternal income indexes jointly. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 8: Effects of SET on mother's income after birth- by component

	Panel A: *Full sample*		
	(1) Sickness benefits	(2) Labor income	(3) Parental benefits
postSET×IVF	-1796.531***	9605.880***	283.077
	(303.427)	(1767.108)	(399.809)
IVF	3738.523***	-2298.335*	681.948**
	(266.353)	(1323.483)	(304.259)
Effect size	-48 %	-418 %	42 %
FDR p-value (Treat)	0.000	0.000	0.566
Bonferroni p-value (Treat)	0.000	0.000	1.000
R-Squared	0.043	0.236	0.170
Observations	936777	936777	936777
Mean of dep. var.	8540	122553	52793
	Panel B: *First-time mothers*		
	(1) Sickness benefits	(2) Labor income	(3) Parental benefits
postSET×IVF	-2103.971***	8533.427***	1928.888***
	(344.714)	(2024.570)	(468.170)
IVF	3660.630***	-355.092	-867.183**
	(298.626)	(1478.004)	(350.345)
Effect size	-57 %	-2403 %	-222 %
FDR p-value (Treat)	0.000	0.000	0.000
Bonferroni p-value (Treat)	0.000	0.001	0.001
R-Squared	.047	.231	.208
Observations	413653	413653	413653
Mean of dep. var.	8061	119938	55765

Notes to Table 8. Refer to notes to Table 5. This table presents identical regression results, however now for each component of the maternal income index. Given the multiplicity of outcome variables considered, both FDR and Bonferroni corrected p-values are reported in table footers. These multiple hypothesis corrections are made considering multiple outcomes in child health, maternal health and maternal income indexes jointly. * p<0.1, ** p<0.05, *** p<0.01.

Table 9: Effects of SET on extensive vs intensive margin earnings

	Full-sample		First time mothers	
	(1)	(2)	(3)	(4)
	Non-zero wage income	Wage income (intensive)	Non-zero wage income	Wage income (intensive)
postSET×IVF	0.004	8955.416***	0.001	8612.008***
	(0.003)	(1792.371)	(0.004)	(2053.151)
IVF	0.003	-3532.617***	0.012***	-2291.831
	(0.003)	(1343.940)	(0.003)	(1496.799)
R-Squared	0.214	0.205	0.191	0.208
Observations	936777	858190	413653	385375
Mean of dep. var.	.91611	133776	.93164	128738
control mean	.91557	133164	.93113	127727
control sd	.27803	105942	.25323	102318

Notes to Table 9. Each column presents a separate OLS regression with DiD estimates of the impact of the SET reform on the average wage income 1-3 year after birth. Columns 1 and 3 present the impact of SET on non-zero income (extensive margin) and columns 2 and 4 is the impact of SET of wage income for women with non-zero earnings (intensive margin). Columns 1-2 present the results for the full sample and columns 3-4 for a sub-sample of first-time mothers. A full set of maternal controls and fixed effects are included in all regressions (as described in Table 3). Standard errors are clustered by mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 10: Sample of mothers with more than one pregnancy: mother fixed effects

	Panel A: *Mother fixed effects excluded*		
	(1)	(2)	(3)
	Child health index	Maternal health index	Maternal labor index
postSET×IVF	0.150***	-0.010	0.118***
	(0.026)	(0.025)	(0.020)
IVF	-0.336***	-0.121***	-0.107***
	(0.020)	(0.017)	(0.016)
Effect size	-45 %	8 %	-110 %
R-Squared	0.014	0.015	0.206
Observations	735771	735771	735165
Mean dep. var.	-0.001	-0.002	0.003
Control mean	0.000	0.000	0.000
Control sd	1.000	1.000	1.000

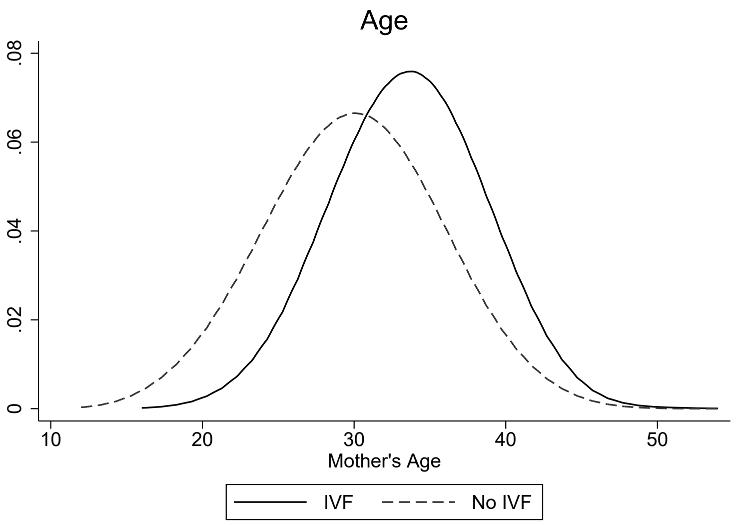
	Panel B: *Mother fixed effects included*		
	(1)	(2)	(3)
	Child health index	Maternal health index	Maternal labor index
postSET×IVF	0.118**	-0.033	0.064**
	(0.059)	(0.057)	(0.028)
IVF	-0.149***	-0.016	-0.073***
	(0.049)	(0.044)	(0.023)
Effect size	-79 %	206 %	-88 %
R-Squared	0.608	0.667	0.896
Observations	735771	735771	735165
Mean dep. var.	-0.001	-0.002	0.003
Control mean	0.000	0.000	0.000
Control sd	1.000	1.000	1.000

Notes to Table 10. The data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007 for a selected sample of mothers with more than one pregnancy. Columns 1-3 present estimates for the full sample and columns 4-6 a sub-sample of first-time mothers. Each column presents a separate OLS regression with DiD estimates of the impact of the SET reform on child health index (column 1), maternal health index (column 2), maternal labor market index (column 3). Panel A presents estimates excluding mother fixed effects and Panel B including mother fixed effects. A full set of maternal controls and fixed effects are included in all regressions (as described in Table 3). Standard errors are clustered by mother. * p<0.1, ** p<0.05, *** p<0.01.

Appendices

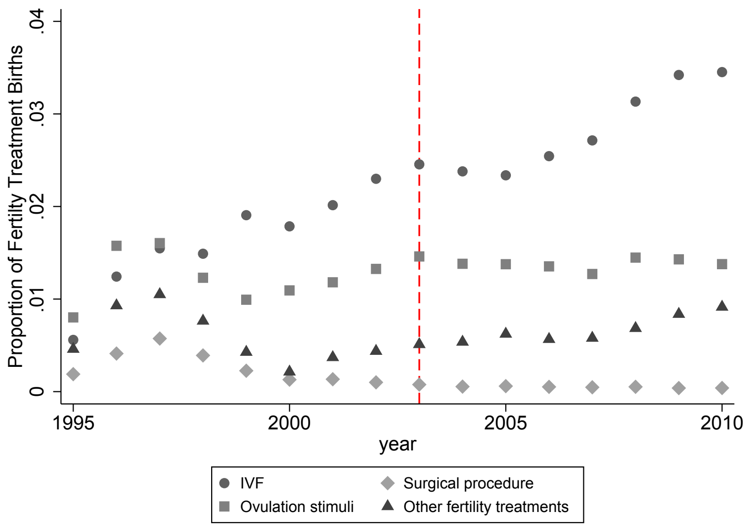
A Appendix Figures and Tables

Figure A1: Age distribution of IVF and non IVF-mothers



The data are obtained from the Swedish Medical Birth Registry for the time period 1998-2007. Figure A1 displays the age distribution among IVF and non-IVF mothers.

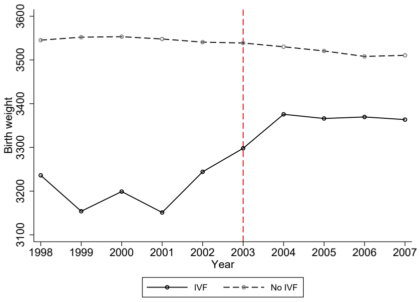
Figure A2: ART treatments



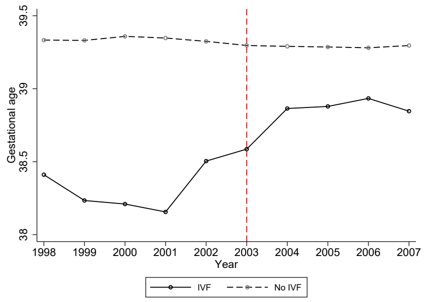
The data are obtained from the Swedish Medical Birth Registry. Trends in different ART treatments are presented in Figure A2. The red-vertical line represents the year of the SET reform.

Figure A3: Child health outcomes (Part 1)

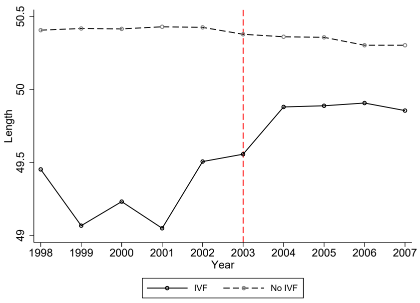
(a) Birth weight



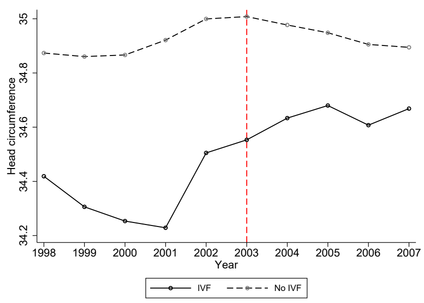
(b) Gestational age



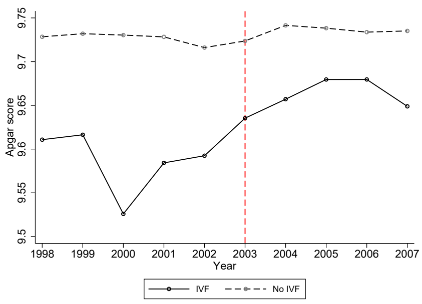
(c) Length



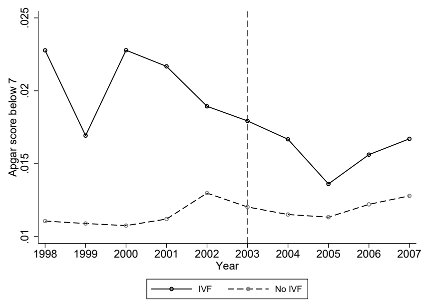
(d) Head circumference



(e) Apgar score

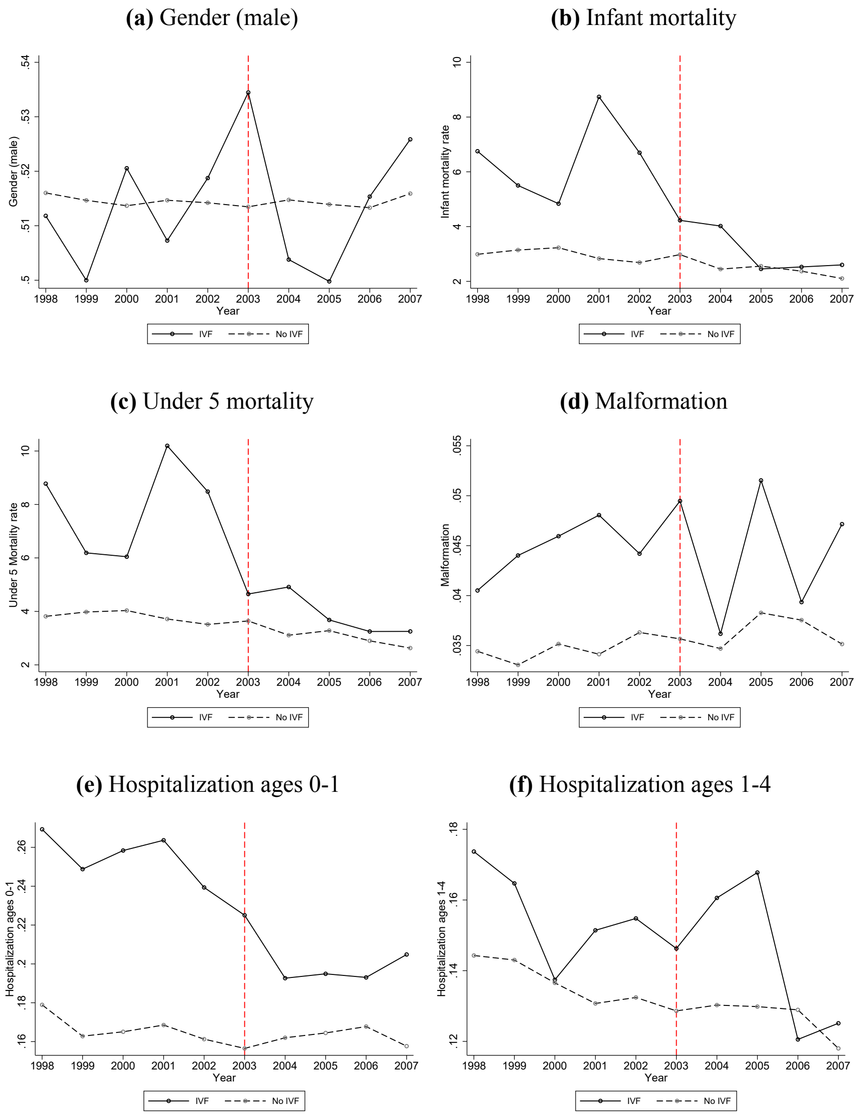


(f) Apgar score below 7



Yearly averages for child health outcomes with and without IVF conception are presented, using microdata obtained from the Swedish Medical Birth Registry and Patient Registry. The red vertical line indicates the year of the SET reform.

Figure A4: Child health outcomes (Part 2)



(a) Gender (male)

(b) Infant mortality

(c) Under 5 mortality

(d) Malformation

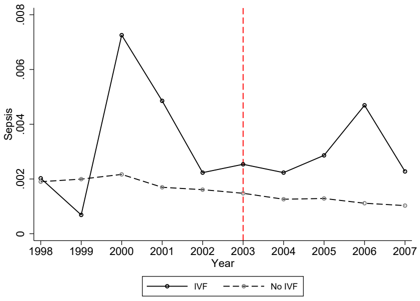
(e) Hospitalization ages 0-1

(f) Hospitalization ages 1-4

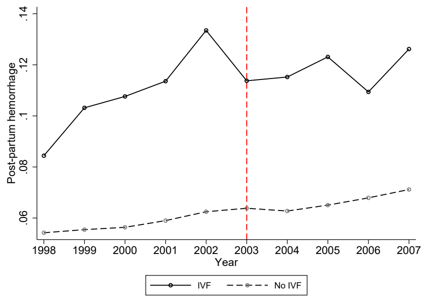
Refer to notes to Appendix Figure A3. Identical plots are presented for additional health outcomes.

Figure A5: Maternal health and labor outcomes

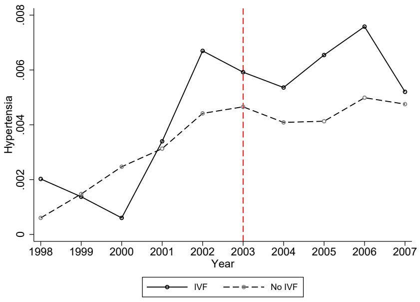
(a) Maternal sepsis



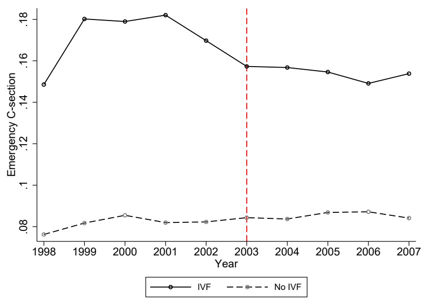
(b) Postpartum hemorrhage



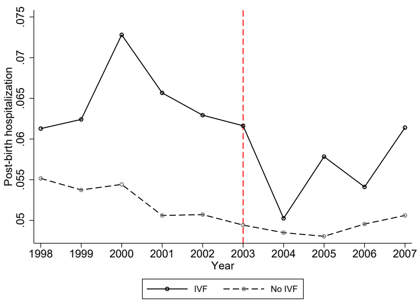
(c) Hypertension



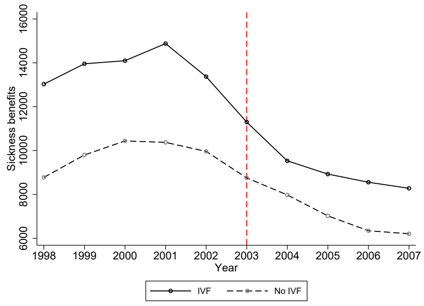
(d) Emergency C-section



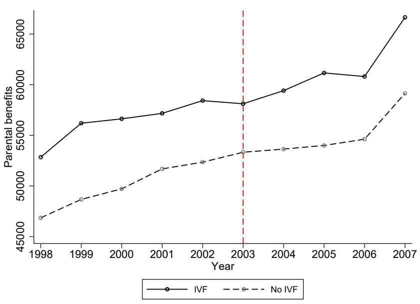
(e) Post-birth hospitalization



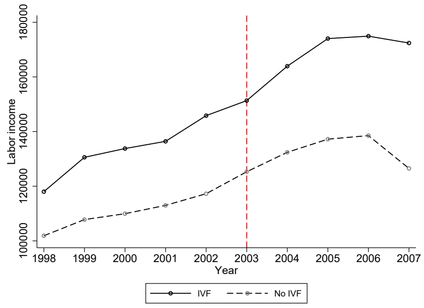
(f) Sickness benefits



(g) Parental benefits



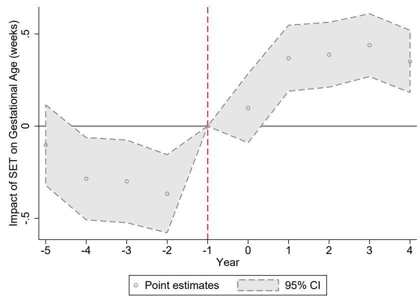
(h) Labor income



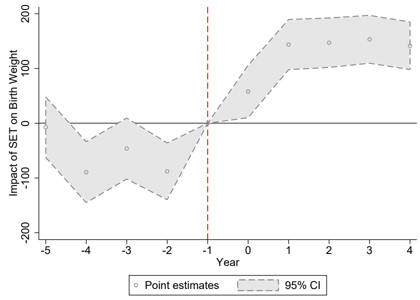
Yearly averages maternal health outcomes with and without IVF conception are presented, using microdata obtained from the from the Swedish Medical Birth Registry and Patient Registry, and Longitudinal integration database for health insurance and labor market studies (LISA). The red vertical line indicates the year of the SET reform.

Figure A6: Child health outcomes

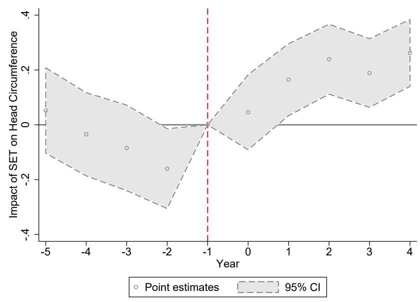
(a) Gestational age



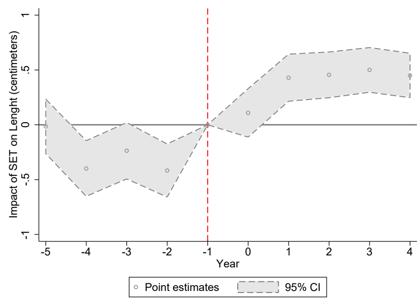
(b) Birth weight



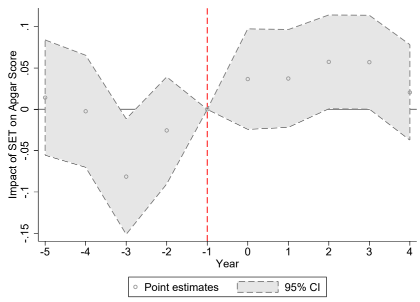
(c) Head circumference



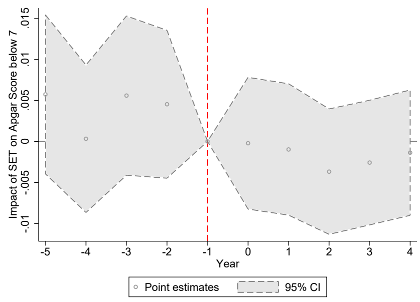
(d) Length



(e) Apgar score

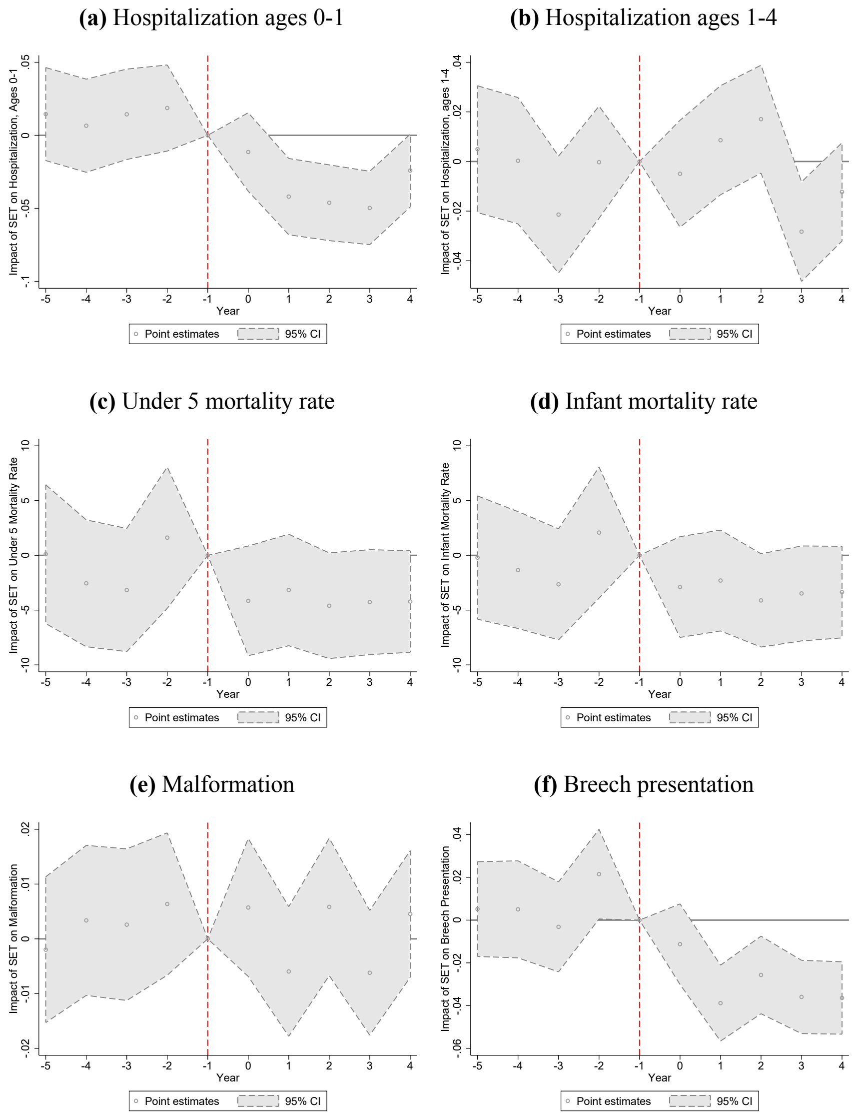


(f) Apgar score below 7



Refer to notes to Figure 6. Identical event study specifications are shown, however now for individual components of the child health index. All details follow those in notes to Figure 6.

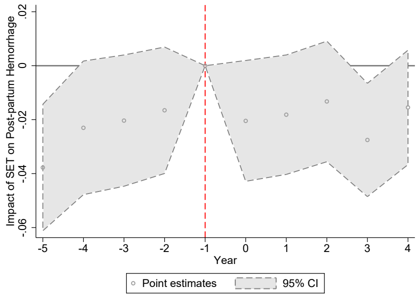
Figure A7: Child health outcomes



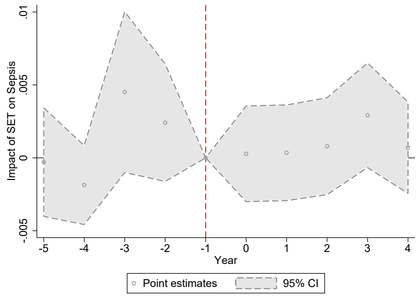
Refer to notes to Figure 6. Identical event study specifications are shown, however now for individual components of the child health index. All details follow those in notes to Figure 6.

Figure A8: Maternal health and labor outcomes

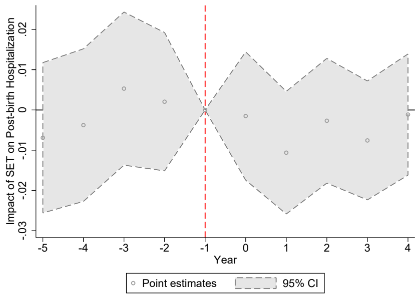
(a) Post-partum hemorrhage



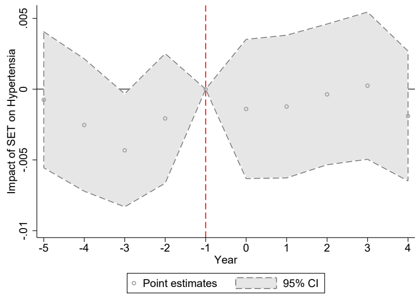
(b) Maternal sepsis



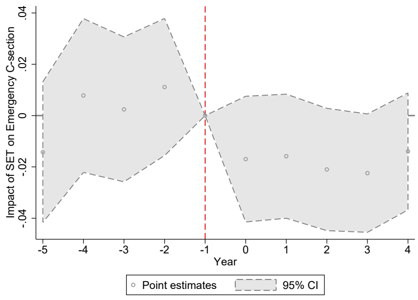
(c) Post birth hospitalization



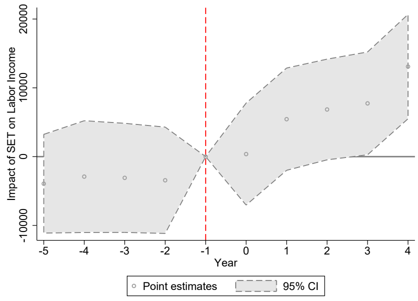
(d) Hypertension



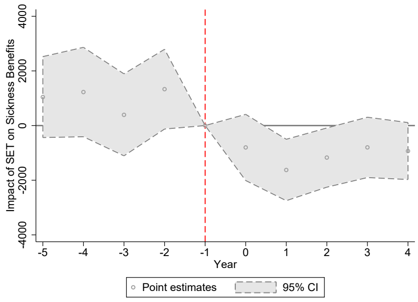
(e) Emergency c-section



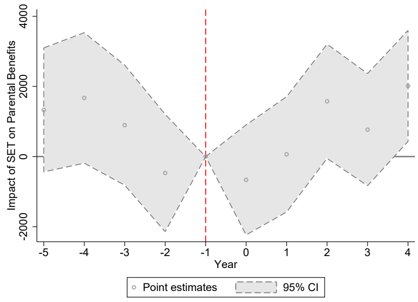
(f) Labor income



(g) Sickness benefits



(h) Parental benefits



Refer to notes to Figure 6. Identical event study specifications are shown, however now for individual components of the maternal health and labour market indexes. All details follow those in notes to Figure 6.

Table A1: Probability of twinning per birth, including trends

	Full sample		*First-time mothers*	
	(1)	(2)	(3)	(4)
	Twin birth	*Twin birth*	*Twin birth*	*Twin birth*
postSET×IVF	-0.132***	-0.129***	-0.131***	-0.129***
	(0.015)	(0.014)	(0.017)	(0.017)
IVF	0.263***	0.249***	0.252***	0.246***
	(0.012)	(0.008)	(0.014)	(0.010)
Mother weight	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Mother height	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Smoking 1st trimester	0.003***	0.003***	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Native	-0.000	-0.000	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
Labor income	0.000	0.000	0.000*	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Sickness benefits	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
IVF specific split time trends	YES	NO	YES	NO
IVF specific global time trends	NO	YES	NO	YES
R-Squared	0.062	0.062	0.068	0.068
Observations	937893	937893	414182	414182
Mean of dep. var.	0.029	0.029	0.029	0.029
Control mean	0.027	0.027	0.027	0.027
Control sd	0.163	0.163	0.162	0.162

Note to Table A1. The data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007. Each column presents a separate OLS regression with DiD estimates of the impact of the SET reform on the probability of twin birth for the full sample (columns 1-2) and first-time mothers (columns 3-4). In columns 1 and 3, an IVF specific split linear time trend is included and in columns 2 and 4, an IVF specific (global) linear time trend is included. A full set of maternal controls and fixed effects are included in all regressions (as described in Table 3). Standard errors are clustered by mother. * p<0.1, ** p<0.05, *** p<0.01.

Table A2: Effects of SET on birth weight and gestational age

Panel A: *Full sample*

	(1) Birth weight <1500 grams	(2) Birth weight <2500 grams	(3) Gestation <28 weeks	(4) Gestation <32 weeks	(5) Gestation <37 weeks
postSET×IVF	-0.012***	-0.068***	-0.005***	-0.013***	-0.083***
	(0.003)	(0.006)	(0.002)	(0.003)	(0.007)
IVF	0.020***	0.112***	0.008***	0.025***	0.157***
	(0.002)	(0.005)	(0.001)	(0.003)	(0.006)
FDR p-value (Treat)	0.000	0.000	0.009	0.000	0.000
Bonferroni p-value (Treat)	0.000	0.000	0.125	0.000	0.000
R-Squared	0.010	0.021	0.012	0.012	0.018
Observations	935714	935714	937893	937893	937893
Mean of dep. var.	0.008	0.042	0.003	0.009	0.113

Panel B: *First-time mothers*

	(1) Birth weight <1500 grams	(2) Birth weight <2500 grams	(3) Gestation <28 weeks	(4) Gestation <32 weeks	(5) Gestation <37 weeks
postSET×IVF	-0.015***	-0.078***	-0.006***	-0.016***	-0.091***
	(0.003)	(0.007)	(0.002)	(0.004)	(0.008)
IVF	0.023***	0.120***	0.010***	0.027***	0.156***
	(0.003)	(0.006)	(0.002)	(0.003)	(0.007)
FDR p-value (Treat)	0.000	0.000	0.009	0.000	0.000
Bonferroni p-value (Treat)	0.000	0.000	0.125	0.000	0.000
R-Squared	0.015	0.024	0.017	0.017	0.022
Observations	413106	413106	414180	414180	414180
Mean of dep. var.	0.010	0.055	0.004	0.012	0.126

Note to Table A2. The data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007. Each column presents a separate OLS regression with DiD estimates of the impact of the SET reform on child health outcomes regarding low birthweight and prematurity. Panel A presents the results for the full sample and Panel B for a sub-sample of first-time mothers. A full set of maternal controls and fixed effects are included in all regressions (as described in Table 3). Given multiplicity of tests, both FDR and Bonferroni corrected p-values are reported in table footers. Standard errors are clustered by mother. * p<0.1, ** p<0.05, *** p<0.01.

Table A3: Effects of SET on child and maternal outcomes, including trends

Panel A: IVF specific split linear time trends	*Full sample*			*First-time mothers*		
	(1) Child health index	(2) Maternal health index	(3) Maternal labor index	(4) Child health index	(5) Maternal health index	(6) Maternal labor index
postSET×IVF	0.137***	0.106***	0.061**	0.097**	0.111**	0.105***
	(0.038)	(0.037)	(0.029)	(0.043)	(0.044)	(0.036)
IVF	-0.357***	-0.224***	-0.131***	-0.292***	-0.217***	-0.159***
	(0.029)	(0.028)	(0.022)	(0.033)	(0.033)	(0.028)
R-Squared	0.019	0.023	0.196	0.021	0.027	0.198
Observations	937893	937893	936777	414180	414180	413652
Mean of dep. var.	-0.003	-0.003	0.005	-0.003	-0.004	0.006
Control mean	0.000	-0.000	-0.000	0.000	0.000	-0.000
Control sd	1.000	1.000	1.000	1.000	1.000	1.000

Panel B: IVF specific linear time trends	*Full sample*			*First-time mothers*		
	(1) Child health index	(2) Maternal health index	(3) Maternal labor index	(4) Child health index	(5) Maternal health index	(6) Maternal labor index
postSET×IVF	0.130***	0.099***	0.054*	0.094**	0.102**	0.099***
	(0.037)	(0.037)	(0.028)	(0.042)	(0.044)	(0.036)
IVF	-0.327***	-0.197***	-0.100***	-0.277***	-0.178***	-0.136***
	(0.022)	(0.022)	(0.017)	(0.024)	(0.025)	(0.021)
R-Squared	0.019	0.023	0.196	0.021	0.027	0.198
Observations	937893	937893	936777	414180	414180	413652
Mean of dep. var.	-0.003	-0.003	0.005	-0.003	-0.004	0.006
Control mean	0.000	-0.000	-0.000	0.000	0.000	-0.000
Control sd	1.000	1.000	1.000	1.000	1.000	1.000

Note to Table A3. The data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007. Columns 1-3 present estimates for the full sample and columns 4-6 a sub-sample of first-time mothers. Each column presents a separate OLS regression with DiD estimates of the impact of the SET reform on child health index (columns 1 and 4), maternal health index (columns 2 and 5), maternal labor market index (columns 3 and 6). Panel A includes IVF specific split linear time trends and Panel B includes IVF specific (global) linear time trends. A full set of maternal controls and fixed effects are included in all regressions (as described in Table 3). Standard errors are clustered by mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A4: Maternal composition

	Panel A: *Full sample*										
	(1) age	(2) weight (kg)	(3) height (cm)	(4) education	(5) labor income	(6) sick benefits	(7) native	(8) smoke	(9) asthma	(10) epilepsy	(11) u. colitis
postSET×IVF	0.190***	-0.086	0.003	0.107***	0.032	-0.005	-0.000	0.008**	0.003	0.000	-0.000
	(0.063)	(0.195)	(0.099)	(0.020)	(0.033)	(0.059)	(0.005)	(0.003)	(0.004)	(0.001)	(0.002)
IVF	4.446***	1.798***	0.799***	0.059***	0.845***	0.814***	0.031***	-0.050***	-0.006*	-0.001	0.005***
	(0.050)	(0.156)	(0.079)	(0.017)	(0.026)	(0.045)	(0.004)	(0.003)	(0.003)	(0.001)	(0.001)
R^2	0.193	0.025	0.010	0.058	0.060	0.122	0.042	0.021	0.009	0.004	0.005
Obs	937,891	835,139	873,962	932,257	928,502	906,708	937,393	880,138	937,893	937,893	937,893
Mean of dep. var.	30.087	67.897	166.417	4.462	10.227	2.870	0.813	0.095	0.068	0.004	0.006

Note to Table A4. The data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007. Panel A presents the results for the full sample and Panel B for a sub-sample of first-time mothers. Each column presents a separate OLS regression with DiD estimates of the impact of the SET reform on maternal age (column 1), weight (column 2), height (column 3), education (column 4), labor income prior birth (column 5), sickness benefits prior birth (column 6), nationality (Swedish) (column 7), smoking (column 8), asthma (column 9), epilepsy (column 10) and ulcerative colitis (column 11). Birth number, county and conception date fixed effects are included in each regression. Standard errors are clustered by mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A5: Effects of SET on child and maternal outcomes, excluding controls

	Full sample			*First-time mothers*		
	(1) Child health index	(2) Maternal health index	(3) Maternal labor index	(4) Child health index	(5) Maternal health index	(6) Maternal labor index
postSET×IVF	0.202***	0.033*	0.118***	0.189***	0.051**	0.157***
	(0.019)	(0.020)	(0.017)	(0.022)	(0.023)	(0.020)
IVF	-0.398***	-0.268***	0.093***	-0.338***	-0.239***	0.021
	(0.016)	(0.015)	(0.013)	(0.018)	(0.018)	(0.016)
R-Squared	0.006	0.006	0.032	0.012	0.011	0.029
Observations	937893	937893	936777	414180	414180	413652
Mean of dep. var.	-0.003	-0.003	0.005	-0.003	-0.004	0.006
Control mean	0.000	-0.000	-0.000	0.000	0.000	-0.000
Control sd	1.000	1.000	1.000	1.000	1.000	1.000
Altonji et al. (2005) Ratio	14.5	32.0	8.8	36.8	-11.2	156.0

Note to Table A5. The data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007. Columns 1-3 present estimates for the full sample and columns 4-6 a sub-sample of first-time mothers. Each column presents a separate OLS regression with DiD estimates of the impact of the SET reform on child health index (columns 1 and 4), maternal health index (columns 2 and 5), maternal labor market index (columns 3 and 6). Maternal and child controls are excluded. Standard errors are clustered by mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A6: Summary statistics, IVF mothers with 1 or more than 1 birth

IVF mothers with:	*≥ 1 birth*	*Only 1 birth*	*Difference*	
	(1) Mean	(2) Mean	(3) T-test	(4) P-values
Twin birth	0.123	0.221	-22.558	0.000
Planned C-section	0.145	0.168	-5.475	0.000
Emergency C-section	0.108	0.204	-22.884	0.000
Maternal sepsis	0.002	0.004	-3.358	0.001
Postpartum hemorrhage	0.091	0.129	-10.318	0.000
Post-birth hospitalization	0.057	0.061	-1.411	0.158
Age	33.552	33.369	3.555	0.000
Weight (kilograms)	167.335	167.358	-0.300	0.764
Height (centimeters)	68.608	68.883	-1.809	0.070
BMI	24.511	24.584	-1.411	0.158
Asthma	0.060	0.074	-4.643	0.000
Ulcerative colitis	0.010	0.011	-0.741	0.459
Epilepsy	0.004	0.005	-1.479	0.139
Hypertensia	0.005	0.005	0.167	0.868
Smoking 1st trimester	0.044	0.041	1.383	0.167
Smoking 3rd trimester	0.028	0.023	2.843	0.004
Education	4.691	4.731	-2.539	0.011
Labor income	58.220	70.097	-24.686	0.000
Sickness benefits	2.041	1.566	7.289	0.000
N births	18334	11154		
N mothers	9931	9831		

Note to Table A6. The data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies. The sample includes IVF mothers for the time period 1998-2007. Mean values, along with t-tests for two-sided tests of equality of means and corresponding p-values for the t-tests are displayed.

Table A7: Effects of SET on child and maternal outcomes, controlling for trends in IVF-specific maternal characteristic

	Full sample			*First-time mothers*		
	(1) Child health index	(2) Maternal health index	(3) Maternal labor index	(4) Child health index	(5) Maternal health index	(6) Maternal labor index
postSET×IVF	0.149***	0.112***	0.059**	0.122***	0.120***	0.105***
	(0.038)	(0.039)	(0.028)	(0.044)	(0.046)	(0.035)
IVF	-1.066***	0.139	-1.133***	-0.927***	0.536*	-0.627**
	(0.270)	(0.282)	(0.228)	(0.309)	(0.320)	(0.270)
R-Squared	0.021	0.027	0.263	0.024	0.033	0.266
Observations	777363	777363	776837	337118	337118	336913
Mean of dep. var.	0.004	-0.001	0.031	0.001	-0.004	0.046
Control mean	0.000	-0.000	-0.000	0.000	-0.000	0.000
Control sd	1.000	1.000	1.000	1.000	1.000	1.000

Note to Table A7. The data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007. Columns 1-3 present estimates for the full sample and columns 4-6 a sub-sample of first-time mothers. Each column presents a separate OLS regression with DiD estimates of the impact of the SET reform on child health index (columns 1 and 4), maternal health index (columns 2 and 5), maternal labor market index (columns 3 and 6). All regressions include IVF-specific maternal characteristic time trends (for education, labor and sickness benefits, nationality, previous health conditions and behavior and age), a full set of fixed effects (conception date, birth order and county) and binary variables for missing values. Standard errors are clustered by mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A8: Robustness: additional sensitivity

	Panel A: *Removing 2001-2002*			
	(1) Twin birth	(2) Child health index	(3) Maternal health index	(4) Maternal labor index
postSET×IVF	-0.168***	0.177***	0.002	0.100***
	(0.010)	(0.024)	(0.024)	(0.020)
IVF	0.267***	-0.342***	-0.135***	-0.121***
	(0.009)	(0.021)	(0.021)	(0.018)
R-Squared	0.055	0.019	0.024	0.197
Observations	754464	754464	754464	753583
Mean of dep. var.	0.028	-0.003	-0.004	0.005
Control mean	-0.000	-0.000	-0.000	-0.000
Control sd	1.000	1.000	1.000	1.000

	Panel B: *Removing region of Skåne*			
	(1) Twin birth	(2) Child health index	(3) Maternal health index	(4) Maternal labor index
postSET×IVF	-0.165***	0.188***	0.027	0.107***
	(0.008)	(0.020)	(0.020)	(0.016)
IVF	0.265***	-0.348***	-0.158***	-0.122***
	(0.007)	(0.016)	(0.016)	(0.013)
R-Squared	0.063	0.019	0.024	0.196
Observations	854191	854191	854191	853191
Mean of dep. var.	0.029	-0.003	-0.003	0.005
Control mean	-0.000	-0.000	-0.000	-0.000
Control sd	1.000	1.000	1.000	1.000

	Panel C: *Removing 2005-2007*			
	(1) Twin birth	(2) Child health index	(3) Maternal health index	(4) Maternal labor index
postSET×IVF	-0.150***	0.154***	0.044*	0.067***
	(0.009)	(0.024)	(0.025)	(0.019)
IVF	0.267***	-0.357***	-0.170***	-0.114***
	(0.006)	(0.016)	(0.015)	(0.013)
R-Squared	0.076	0.021	0.023	0.186
Observations	631952	631952	631952	631184
Mean of dep. var.	0.029	-0.002	-0.002	0.002
Control mean	0.000	0.000	0.000	0.000
Control sd	1.000	1.000	1.000	1.000

Note to Table A8. The data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007. Each column presents a separate OLS regression with DiD estimates of the impact of the SET reform on the probability of twin birth (column 1), child health index (column 2), maternal health index (column 3), and maternal labor market index (column 4). In Panel A, the time period 2001-2002 is omitted. In Panel B, the region of Skåne is omitted and in Panel C, the time period 2005-2007 is omitted. A full set of maternal controls and fixed effects are included in all regressions (as described in Table 3). Standard errors are clustered by mother. * p<0.1, ** p<0.05, *** p<0.01.

Table A9: Effects of SET on child and maternal outcomes, 2SLS estimates

	Full sample				*First-time mothers*			
	First stage	*Second stage*			*First stage*	*Second stage*		
	(1) Singleton	(2) Child health index	(3) Maternal health index	(4) Maternal labor index	(5) Singleton	(6) Child health index	(7) Maternal health index	(8) Maternal labor index
postSET×IVF	0.168***				0.177***			
	(0.007)				(0.009)			
Singleton		1.122***	0.186	0.630***		1.022***	0.308**	0.868***
		(0.106)	(0.115)	(0.095)		(0.115)	(0.126)	(0.113)
IVF	-0.268***	-0.055***	-0.120***	0.041**	-0.268***	-0.040**	-0.076***	0.071***
	(0.006)	(0.019)	(0.022)	(0.017)	(0.007)	(0.020)	(0.023)	(0.020)
F-statistic	515.2				413.4			
Observations	937893	937893	937893	936777	414182	414182	414182	413654
Mean of dep. var.	0.971	0.000	-0.000	-0.000	0.971	-0.000	0.000	0.000
Control mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Control sd	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note to Table A9. The data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007. Column 1 presents an OLS regression of the probability of having a singleton birth after SET for the full sample and column 5 the probability for first-time mothers. Columns 2-4 present estimates for the full sample and columns 6-8 a sub-sample of first-time mothers. Each column presents a separate 2SLS regression of the impact of having a singleton birth due to SET on child health (columns 2 and 6), maternal health (columns 3 and 7) and maternal income after birth (columns 4 and 8). A full set of maternal controls and fixed effects are included in all regressions (as described in Table 3). Standard errors are clustered by mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A10: Heterogeneous effects of SET on child health index

Dependent variable: Child health index								
Panel A Sub-Sample:	Birth order			Age groups				
	(1) Birth order 1	(2) Birth order 2	(3) Birth order ≥ 3	(4) Ages < 25	(5) Ages 25-29	(6) Ages 30-34	(7) Ages 35-38	(8) Ages ≥ 39
postSET×IVF	0.200***	0.185***	0.0460	0.309	0.259***	0.242***	0.100***	0.118
	(0.0233)	(0.0391)	(0.0928)	(0.1884)	(0.0494)	(0.0304)	(0.0323)	(0.0750)
IVF	-0.325***	-0.345***	-0.342***	-0.470***	-0.449***	-0.374***	-0.241***	-0.192***
	(0.0193)	(0.0327)	(0.0759)	(0.1571)	(0.0397)	(0.0257)	(0.0259)	(0.0648)
Observations	414180	341726	181347	132592	290706	331919	153399	29275
R^2	0.015	0.016	0.026	0.032	0.020	0.024	0.037	0.145
Mean of dep. var.	-0.0660	0.0575	0.0342	-0.0393	-0.00139	0.0161	-0.00258	-0.0578
Panel B Sub-Sample:	Education			BMI classification				
	(1) Primary education	(2) Secondary education	(3) Tertiary education	(4) Under weight BMI < 18.5	(5) Normal weight BMI 18.5-24	(6) Overweight BMI 25-29	(7) Obesity BMI ≥ 30	
postSET×IVF	0.193***	0.170***	0.198***	-0.0255	0.214***	0.184***	0.119*	
	(0.0441)	(0.0308)	(0.0315)	(0.1230)	(0.0260)	(0.0397)	(0.0640)	
IVF	-0.366***	-0.311***	-0.340***	-0.150	-0.351***	-0.378***	-0.307***	
	(0.0332)	(0.0255)	(0.0271)	(0.0928)	(0.0224)	(0.0324)	(0.0507)	
Observations	236541	383284	312432	19433	498958	211379	90396	
R^2	0.025	0.020	0.023	0.204	0.017	0.028	0.053	
Mean of dep. var.	-0.0561	-0.000487	0.0362	-0.0574	0.0146	0.00274	-0.0484	

Note to Table A10. The data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007. Each column presents a separate OLS regression with DiD estimates of the impact of the SET reform.

Table A11: Heterogeneous effects of SET on maternal health index

Dependent variable: Maternal health index

Panel A Sub-Sample:	Birth order			Age groups				
	(1) Birth order 1	(2) Birth order 2	(3) Birth order $\geq$ 3	(4) Ages < 25	(5) Ages 25-29	(6) Ages 30-34	(7) Ages 35-38	(8) Ages $\geq$ 39
postSET×IVF	0.0674***	-0.0392	-0.0746	-0.0409	0.0761	0.0240	0.0397	0.0557
	(0.0245)	(0.0343)	(0.0737)	(0.1178)	(0.0467)	(0.0273)	(0.0366)	(0.0943)
IVF	-0.179***	-0.103***	-0.0572	-0.157**	-0.241***	-0.125***	-0.162***	-0.160**
	(0.0194)	(0.0239)	(0.0505)	(0.0795)	(0.0357)	(0.0211)	(0.0288)	(0.0762)
Observations	414180	341726	181347	132592	290706	331919	153399	29275
R^2	0.018	0.016	0.025	0.033	0.020	0.023	0.038	0.133
Mean of dep. var.	-0.0825	0.0616	0.0568	0.0383	0.0235	-0.00585	-0.0548	-0.161

Panel B Sub-Sample:	Education			BMI classification			
	(1) Primary education	(2) Secondary education	(3) Tertiary education	(4) Under weight BMI < 18.5	(5) Normal weight BMI 18.5-24	(6) Overweight BMI 25-29	(7) Obesity BMI $\geq$ 30
postSET×IVF	0.0172	0.0206	0.0552*	0.0170	0.0715***	0.0229	-0.0195
	(0.0423)	(0.0310)	(0.0325)	(0.1210)	(0.0239)	(0.0424)	(0.0836)
IVF	-0.152***	-0.160***	-0.188***	-0.224***	-0.193***	-0.164***	-0.0880
	(0.0296)	(0.0232)	(0.0270)	(0.0850)	(0.0189)	(0.0335)	(0.0676)
Observations	236541	383284	312432	19433	498958	211379	90396
R^2	0.027	0.022	0.026	0.205	0.019	0.033	0.058
Mean of dep. var.	-0.0224	-0.00122	0.00849	0.0694	0.0460	-0.0381	-0.187

Note to Table A11. The data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007. Each column presents a separate OLS regression with DiD estimates of the impact of the SET reform.

Table A12: Heterogeneous effects of SET on maternal income index

Dependent variable: Maternal labor index								
Panel A Sub-Sample:	Birth order			Age groups				
	(1) Birth order 1	(2) Birth order 2	(3) Birth order ≥ 3	(4) Ages < 25	(5) Ages 25-29	(6) Ages 30-34	(7) Ages 35-38	(8) Ages ≥ 39
postSET×IVF	0.168***	0.176***	0.0874	0.0253	0.0998***	0.0806***	0.0595**	-0.00880
	(0.0188)	(0.0333)	(0.0890)	(0.0854)	(0.0355)	(0.0240)	(0.0301)	(0.0824)
IVF	-0.172***	-0.130***	-0.214***	-0.196***	-0.262***	-0.179***	0.00180	0.172**
	(0.0150)	(0.0269)	(0.0697)	(0.0688)	(0.0288)	(0.0196)	(0.0245)	(0.0676)
Observations	413652	341330	181156	132462	290390	331500	153184	29241
R^2	0.142	0.112	0.101	0.103	0.090	0.103	0.106	0.193
Mean of dep. var.	0.0726	0.0357	-0.208	-0.442	-0.0623	0.169	0.158	0.0232
Panel B Sub-Sample:	Education			BMI classification				
	(1) Primary education	(2) Secondary education	(3) Tertiary education	(4) Under weight BMI < 18.5	(5) Normal weight BMI 18.5-24	(6) Overweight BMI 25-29	(7) Obesity BMI ≥ 30	
postSET×IVF	0.0837**	0.104***	0.102***	0.185	0.123***	0.164***	0.168***	
	(0.0339)	(0.0239)	(0.0271)	(0.1448)	(0.0213)	(0.0320)	(0.0579)	
IVF	-0.0957***	-0.126***	-0.134***	-0.0969	-0.157***	-0.168***	-0.229***	
	(0.0253)	(0.0197)	(0.0227)	(0.0946)	(0.0173)	(0.0262)	(0.0488)	
Observations	236388	382984	312023	19393	498400	211173	90344	
R^2	0.069	0.087	0.112	0.297	0.136	0.109	0.107	
Mean of dep. var.	-0.410	-0.0327	0.375	-0.208	0.0859	-0.0578	-0.242	

Note to Table A12. The data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007. Each column presents a separate OLS regression with DiD estimates of the impact of the SET reform.

Table A13: Gelbach decomposition

	(1) Child health index	(2) Maternal health index	(3) Wage income	(4) Decomposition
postSET×IVF	$\Gamma^{childhealth} = 0.188$***	$\Gamma^{momhealth} = 0.032$**	9605.951***	
	(0.014)	(0.014)	(1767.107)	
$\beta^{childhealth}_{wage}$ = child health index			1595.830***	
			(103.322)	
$\beta^{momhealth}_{wage}$ = mom health index			1363.807***	
			(105.794)	
$\Gamma^{childhealth}_{wage} \times \beta^{childhealth}_{wage}$				$\delta^{childhealth}_{wage} = 299.553$***
				(36.700)
$\Gamma^{momhealth}_{wage} \times \beta^{momhealth}_{wage}$				$\delta^{momhealth}_{wage} = 43.831$
				(27.021)
Total explained difference				343.384***
				(47.140)

Note to Table A13. The data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007. A full set of maternal controls and fixed effects are included in all regressions (as described in Table 3). Standard errors are clustered by mother. * p<0.1, ** p<0.05, *** p<0.01. Γ represents each estimate of the SET reform (postSET×IVF) for each potential mechanism as the outcome variable. β indicates the estimate of the potential mechanisms as explanatory variables in the full specification with maternal labor as the outcome variable. The conditional contribution of each component is given by δ, which is computed by multiplying Γ with β.

Table A14: Impact on proportion of IVF births, deliveries per transfer and number of IVF treatments

	(1) Proportion of IVF births	(2) Delivery rate	(3) Started IVF cycles
postSET	-0.002*	0.000	120.733
	(0.001)	(0.009)	(311.976)
Trend	0.001***	-0.000	813.327***
	(0.000)	(0.001)	(60.911)
R^2	0.847	0.056	0.992
Observations	11	11	11
Mean of dep. var.	0.029	0.244	11975.636

Note to Table A14. Aggregate data on proportion of IVF births, deliveries per transfer and number of IVF treatments are collected from annual reports by the Swedish National Board of Health and Welfare for the time period 1998-2008. Each column presents a separate OLS regression with the impact of the SET reform on the proportion of IVF births (column 1), deliveries per transfer (column 2), and number of treatments (column 3). All regressions include a linear time trend. Robust standard errors in parenthesis. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

B Measurement of IVF usage

A number of methodologies exist to consider mis-reporting of treatment variables (Horowitz and Manski, 1995), or selection into treatment (Alderman et al., 2011; Lee, 2009). The case we are concerned with is relatively simple, as we are concerned only with a mis-classification of treated units to be included as part of the control group. Given our application, in general, we are likely to under-estimate the effect size by a small amount. To see why, we provide some simple algebra considering the difference between a DiD estimator where all treated units are correctly classified: $\widehat{\beta}_1$, and an estimator where some portion of treated units are mis-classified as controls $\widehat{\bar{\beta}}_1$. These estimators can, respectively, be written as:

$$\widehat{\beta}_1 = (\bar{Y}_{T1} - \bar{Y}_{C1}) - (\bar{Y}_{T0} - \bar{Y}_{C0}),$$

where $\bar{Y}_{T1}$ refers to average outcomes among treated following treatment, $\bar{Y}_{C1}$ refers to average outcomes among controls following treatment, and $\bar{Y}_{T0}$ and $\bar{Y}_{C0}$ are the same values prior to treatment. The biased estimator, on the other hand, is:

$$\widehat{\bar{\beta}}_1 = (\bar{Y}_{T1} - \bar{\bar{Y}}_{C1}) - (\bar{Y}_{T0} - \bar{\bar{Y}}_{C0}),$$

where now $\bar{\bar{Y}}_{C1}$ includes a small portion of the incorrectly classified treated units, and similarly for $\bar{\bar{Y}}_{C0}$. In particular,

$$\bar{\bar{Y}}_{C1} = \frac{T_{C1}}{T_{C1} + T_{mc^1}} \bar{Y}_{C1} + \frac{T_{mc^1}}{T_{C1} + T_{mc^1}} \bar{Y}_{T1}.$$

Here T_{C1} refers to the total number of control units in period 1, and T^1_{mc} refers to the total number of mis-classified treated units included as controls following treatments. A similar value is defined for $\bar{\bar{Y}}_{C0}$. It is worth noting here that $\bar{\bar{Y}}_{C1}$ will equal the true value $\bar{Y}_{C1}$ in two circumstances: either if T^1_{mc} is zero (and there is no mis-classification), or if $\bar{Y}_{C1} = \bar{Y}_{T1}$ and so mis-classification does not matter. Now, we can calculate the bias in the diff-in-diff estimate as the difference between the true value $\widehat{\beta}_1$ and the observed value with misclassification $\widehat{\bar{\beta}}_1$. This is calculated as:

$$\begin{aligned}
Bias(\widehat{\beta}_1) &= \widehat{\beta}_1 - \widehat{\bar{\beta}}_1 = (\bar{\bar{Y}}_{C1} - \bar{Y}_{C1}) - (\bar{\bar{Y}}_{C0} - \bar{Y}_{C0}) \\
&= \left(\frac{T_{C1}}{T_{C1} + T^1_{mc}} \bar{Y}_{C1} + \frac{T^1_{mc}}{T_{C1} + T^1_{mc}} \bar{Y}_{T1} - \bar{Y}_{C1} \right) - \\
&\quad \left(\frac{T_{C0}}{T_{C0} + T^0_{mc}} \bar{Y}_{C0} + \frac{T^0_{mc}}{T_{C0} + T^0_{mc}} \bar{Y}_{T0} - \bar{Y}_{C0} \right) \\
&= \left(\frac{T^1_{mc}}{T_{C1} + T^1_{mc}} \bar{Y}_{T1} - \frac{T^1_{mc}}{T_{C1} + T^1_{mc}} \bar{Y}_{C1} \right) - \\
&\quad \left(\frac{T^0_{mc}}{T_{C0} + T^0_{mc}} \bar{Y}_{T0} - \frac{T^0_{mc}}{T_{C0} + T^0_{mc}} \bar{Y}_{C0} \right)
\end{aligned} \tag{4}$$

If we are further willing to assume that the misclassification of treatment units is constant over time (in our setting, that IVF births are constantly under-reported by 30%), this can be further simplified to:

$$Bias(\widehat{\beta}_1) = \frac{T_{mc}}{T_C + T_{mc}} [(\bar{Y}_{T1} - \bar{Y}_{C1}) - (\bar{Y}_{T0} - \bar{Y}_{C0})]. \tag{5}$$

This simple bias formula thus suggests that misclassification will bias the estimate by the true diff-in-diff estimate, scaled by a parameter capturing the degree of mis-classification of the control group. In our case, given that this proportion $\frac{T_{mc}}{T_C+T_{mc}}$ is small, biases in estimates will also be small. And indeed, we can provide a back-of-the-envelope calculation of this bias using the observed values in the data. Assuming that the proportion of mis-classified IVF births is constant over time, we have that $\frac{T_{mc}}{T_C+T_{mc}} = \frac{9{,}336}{916{,}110} = 0.0102$. Now, for the case of birth weight, we can approximate the bias using values from the data as:

$$\begin{aligned} Bias(\widehat{\beta}_1^{BW}) &= \frac{T_{mc}}{T_C + T_{mc}}[(\bar{Y}_{T1} - \bar{Y}_{C1}) - (\bar{Y}_{T0} - \bar{Y}_{C0})] \\ &= 0.0102 \times [(3200 - 3550) - (3400 - 3530)] = -2.244 \end{aligned} \tag{6}$$

In this case, we estimate that the bias in the estimate of SET is likely to be around 2 or 3 grams. When compared to the original estimate from table 8 of 176 grams, we see that this suggests a (relatively) quite small attenuation of estimated effects.