Multiple Births, Birth Quality and Maternal Labor Supply: analysis of IVF reform in sweden

Sonia Bhalotra[1], Damian Clarke[2], Hanna Mühlrad[3], and Mårten Palme[4]

[1]University of Essex
[2]Universidad de Santiago de Chile
[3]Lund University
[4]Stockholm Univeristy

November 14, 2018

Abstract

In this study we examine the passage of a reform to in-vitro fertilization (IVF) procedures in Sweden in 2003. Following publication of medical evidence showing that pregnancy success rates could be maintained using single rather than multiple embryo transfers, the single embryo transfer (SET) was mandated as the default IVF procedure. Using linked registry data for the period 1998-2007, we find that the SET reform was associated with a precipitous drop in the share of multiple births of 63%. This narrowed differences in health between IVF and non-IVF births by 53%, and differences in the labor market outcomes of mothers three years after birth by 85%. For first time mothers it also narrowed the gap in maternal health between IVF and non-IVF births by 36%. Our findings imply that more widespread adoption of SET could lead to massive gains, reducing hospitalization costs and the foregone income of mothers and improving the long-run socioeconomic outcomes of children. This is important given that the share of IVF facilitated births exceeds 3% in several industrialized countries and is on the rise.

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

[*] We are very grateful to Andreea Mitrut, Randi Hjalmarsson, Mircea Trandafir, Sverker Ek, and seminar audiences at the Department of Economics at University of Gothenburg and ISER, 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 as well as 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; Abramowitz, J., 2017; Kroeger and La Mattina, 2017; 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).

However, there are substantial costs associated with IVF. In addition to costs of the procedure, estimated to range from 40,000 to over 200,000 USD per IVF birth in the US (Bitler, 2008), there are costs arising from adverse pregnancy outcomes (Sazonova et al., 2011) and adverse birth outcomes (Kalra and Barnhart, 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, be presented in breech position, have low birth weight and have lower Apgar scores at birth. This implies additional costs of neonatal and maternal health care that are potentially large (Almond et al., 2010) and, over and above, with expected long-run costs in terms of lower cognitive skills, educational attainment, income and life expectancy among IVF births (Behrman and Rosenzweig, 2004; Bhalotra et al., 2017; Bharadwaj et al., 2013; Black et al., 2007; Oreopoulos et al., 2008).

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; 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 to approximately 1.6% among non-IVF pregnancies (Maheshwari et al., 2011). The reason that IVF births are so much more likely to be multiple births is that IVF has typically involved multiple embryo transfers to increase the chances of success. 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 two embryos (see section 2.3). In light of evidence of this in medical studies, in January 2003, the Swedish National Board of Health and Welfare mandated a SET as the default IVF procedure. Exceptions were allowed as detailed in section 2.3.

Sweden was the pioneer. In July 2003, Belgium followed suit. In 2010, Turkey and Quebec implemented a similar reform. However, as we write in 2017, most IVF treatments in the United States, the United Kingdom and other countries continue to involve a double embryo transfer (DET). Using population registers for Sweden for 1998-2007, we compare outcomes for IVF vs non-IVF births before and after the SET reform, tracing impacts of the reform on indices of child and maternal health and maternal labor market outcomes in the years following birth. We document a post-reform drop in the probability that an IVF birth is a twin of 63%. We identify significant improvements in child health and maternal labor market outcomes overall and, among first-time mothers, also significant improvements in maternal health. We estimate that the reform narrowed the gap between IVF and non-IVF births by 53% in an index of child health and 85% in an index of maternal labor market outcomes. Among first-time mothers, the child health gap narrows by 58%, the maternal labor outcomes gap by 96% and the gap in an index of maternal health by 36%. The increase in child health is evident in indicators of child health, including fetal growth indicated by birth weight, length, head circumference, longer gestational age and fewer complications such as breech presentation, emergency C-section and hospitalization. The improvement in maternal labor markets outcomes is driven by higher labor incomes within three years of birth, and lower sickness benefits.[1] These are intent to treat estimates since there was not perfect compliance with SET. We adjust for selection into IVF and, conditional on IVF, for selection into SET. We allow that omitted trends are different for IVF vs non-IVF births. Since we have many outcomes, we check robustness of the estimates to adjusting for multiple hypothesis testing.

The documented improvements in maternal and child health flow directly from the reduction in the share of multiple births, consistent with the evidence cited above. The improvements in maternal labor supply are likely to flow from both the direct effect of

[1] This is a marker of lower sickness so although we do not find no average impacts on the maternal health index, we note that this is some evidence of improved health.

an increase in the share of uniparous pregnancies, and the associated improvements in maternal and newborn health. Evidence of the impacts of fertility on the labor market outcomes of mothers is mixed (see, for example, (Adda et al., 2017; Browning, 1992; Lundborg et al., 2014). Among reasons for the evidence being ambiguous are that extensive margin fertility tends to have larger impacts on wages (Lundborg et al., 2014), that many studies use the occurrence of a twin birth as an instrument for fertility but twin births occur disproportionately to healthier women who may have unobservable characteristics that predispose them towards stronger labor market performance (Bhalotra and Clarke, 2016), or that the setting matters and the opportunity cost of women's time varies with the level of economic development (Aaronson et al., 2017).[2] Studies that instrument fertility with twins effectively model the impact of one unexpected child at any parity, while Lundborg et al. (2014) leverage quasi-random variation in IVF success rates so they effectively model the impact of an IVF birth at first parity and as this is often a twin, often this involves the number of children jumping from 0 to 2. Like (Lundborg et al., 2014) we study an IVF sample but our quasi-experiment (the SET reform) effectively delivers a discontinuous change in the number of children in the opposite direction, from 2 to 1. Although this is not discussed, previous studies are effectively modeling not only the impact of an increase in the number of births but also the occurrence of twin births which we know are less healthy. Similarly, the impact of a decrease in fertility in our study is combined with impacts of an improvement in the health of the child. Using the procedure in Gelbach (2016), we estimate that only a small portion of the improvement in the mother's labor market outcomes can be attributed to improvements in maternal and child health.

The Swedish SET reform has been analyzed in the biomedical literature but many of these studies suffer from small samples and/or are unable to identify causal effects of SET (Karlström and Bergh, 2007; Lundin and Bergh, 2007; Saldeen and Sundström, 2005; Sazonova et al., 2011; Thurin et al., 2004). The closest relatives of our study in the economics literature are Bitler (2008) and Lundborg et al. (2014). Bitler (2008) analyzes the negative impact of infertility treatment mandates in the United States on birth outcomes, underlining the costs of the rising share of twin births. Lundborg et al. (2014), discussed above, examine the impact of fertility among IVF users on female labor supply in Den-

[2] This is a large literature and there are other variations including whether the outcome of interest is labor supply or earnings, and the horizon over which it is measured.

mark. Our paper is similar insofar as it investigates variation in fertility brought about by IVF technologies, but we consider variation in the *number* of births subject to successful IVF treatment, while they study variation in the success of IVF treatment. We gain exogenous variation in the number of births in the IVF sample from introduction of the SET reform, while they demonstrate that variation in IVF success rates is orthogonal to the observed characteristics of women. Our purpose is to analyze thoroughly the impacts of the SET mandate on health and labor market outcomes, contributing to the case for its more widespread adoption while theirs is to re-visit the classical question of whether fertility influences women's labor supply.

The findings of this study have important implications for other countries considering policy reform similar to that initiated by Sweden in 2003. While there has been a shift from DET to SET in multiple countries including Scandinavia and Belgium, and Turkey, DET or higher order 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). There appears to be a lack of information on the advantages of SET among couples seeking IVF treatments and this may be a function of the financial incentives of insurance companies (Pinckney-Clark et al., 2016). Countries with mostly private funding and/or insurance systems appear to have a harder time implementing SET (Karlström and Bergh, 2007). Our findings shed light on the gains from SET not only in terms of child health but also in terms of labor market outcomes for the mother.

This paper is structured in the following way. Section 2 presents a description of IVF in a Swedish context and the implementation of the SET policy. In Sections 3 and 4, respectively, the data and empirical strategy are described. Section 5 presents the results. In Section 6 the findings are discussed, concluding the study.

2 Background

2.1 IVF treatments in Sweden

In Sweden, all residents (registered in the population registry) have access to heavily subsidized health care provided by both private and public health care providers and

IVF treatments are covered under certain conditions discussed below.[3] Health care in Sweden is mainly funded by tax revenues and only 2% of all residents have private health insurance (Anell, 2008). IVF procedures are primarily regulated under the law on genetic integrity.[4] Access and funding varies across counties, with local county council boards being responsible for the setting and implementation of rules and requirements of IVF in their jurisdiction (Vårdguiden, 2017). For example, the maximum maternal age for government funded IVF differs across regions in the country. To take two cases, in Örebro county the upper age limit is 43 while the limit is 37 in Norrbotten county (Alm, 2010). The provision of IVF is allowed in private and public regimes, subject to the approval of the Health and Social Care Inspectorate (Inspektionen för vård och omsorg, IVO).[5]

The Swedish Association of Local Authorities and Regions provides guidelines for IVF treatments, including eligibility guidance although the local health care provider is responsible for making sure that these requirements are met and enforcement is not strict (SKL, 2016). The criteria laid down are as follows. First, the couple undergoing treatment should be in a stable union, either legally married or co-habitating for at least two years, although starting in 2016 single women are also allowed to access publicly funded IVF treatment. Second, the woman should have no previous children, either biological or adopted. 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, and that the county specific age restriction is met. Maternal age for starting the first treatment should be below age 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. Fourth, three rounds of treatment (follicle aspiration) should be offered to each couple, and any remaining embryos and eggs of good quality should be frozen. Finally, additional conditions including mental and physical illness or disability are to be considered before offering treatment.

[3] For most medical services, there is a small fee until the patient reaches the maximum amount of 1100 SEK (approximately 110 USD) annually. Above this fee the health care services have usually no additional costs.

[4] 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 in other laws, including the Children and Parents' Code (Föräldrabalk (1949:381)) and the Health and Medical Services Act (Hälso-och sjukvårdslag (1982:763)).

[5] 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.

2.2 Number of embryo transfers and pregnancy success

During the 1980s and early 1990s, IVF had relatively low delivery rates per treatment. Therefore, multiple embryos were usually transferred in order to maximize the probability of a successful pregnancy (Karlström and Bergh, 2007). While this improved success rates it also raised the share of multiple births among IVF relative to non-IVF births. In response to multiple births exhibiting worse neonatal outcomes (Bergh et al., 1999), in the early 1990s Swedish clinics implemented a voluntary shift from triple embryo transfer (TET) to double embryo transfer (DET). This reduced triple births drastically towards zero without lowering the delivery rate but there remained a high prevalence of twins among IVF births throughout the 1990s (Karlström and Bergh, 2007). It was not until 2003 that the single embryo transfer (SET) was introduced as a mandate.

The implementation of SET was a response to medical evidence that pregnancy success rates of IVF could be maintained with SET. The pioneering study was by Vilska et al. (1999), who looked at elective SET cases in Finland. Their evidence was broadly supported by one of the largest randomized control trials in this domain, with over 660 participants, conducted at multiple centers in Scandinavia (Thurin et al., 2004). 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 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. Other randomized control trials with smaller samples, and subsequent observational studies provided broadly similar results (Criniti et al., 2005; Gerris et al., 2001; Karlström and Bergh, 2007; Lukassen et al., 2005; Lundin and Bergh, 2007).

2.3 The SET reform

On January 1 2003, the Swedish National Board of Health and Welfare issued new provisions and general guidelines for IVF, mandating that the routine procedure should be to transfer one embryo at a time in IVF treatments, with exceptions allowed 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 double embryo transfer, provided that they were informed about the potential risks for the mother and child of undergoing a DET (Saldeen and Sundström, 2005).[6] The SET reform was motivated to improve birth outcomes and lower costs of neonatal care. Previous studies of SET suggested it achieved its goal of lowering costs per birth. One study estimates that costs six months following birth fell from approximately 160,000 to 90,000 Euros (Thurin et al., 2004) and another estimated that reduced maternal and neonatal hospital stays saved 10,000 Euros per birth (Lukassen et al., 2005).

Although exceptions were permitted, the reform generated a sharp increase in the share of IVF treatments that involved a single embryo transfer, from 30% to 70% within 24 months (see Figure 1a). The pregnancy success rate among IVF users was maintained at about one-quarter (Karlström and Bergh, 2007); also see Figure 2a. The number of IVF treatments performed is smooth around the cut-off (Figure 2b), although Figure 1b shows a slight decrease in the proportion of IVF births following the reform.[7] We shall formally test for discontinuities in the proportion of IVF births and deliveries per transfer (success rates) (see Section 5.2). Over this period there were no other changes in the IVF treatment procedure with respect to medication, technique or equipment (Saldeen and Sundström, 2005).[8]

3 Data and descriptive statistics

3.1 Data

We use Swedish administrative data to examine the impact of the SET reform. In particular, we use the Swedish Multi-Generational Registry (Flergenerationsregistret) provided by Statistics Sweden, which contains all registered people in Sweden after 1961. Our baseline sample consists of cohorts born between 1940 and 1985 including their children and parents. Based on these individuals, we select all women giving birth during 1998-2007

[6]See also the 2003 provisions and general guidelines for IVF from the Swedish National Board of Health and Welfare: *Socialstyrelsens föreskrifter och allmänna råd om assisterad befruktning, SOSFS 2002:13.*

[7]See Figure A1 which shows a gradually rising trend in the proportion of IVF births and the share of twin births in all (IVF and other) births. Trends in the proportions of each type of ART procedures are presented in Figure A1, which shows that IVF is the only ART procedure exhibiting a strong trend.

[8]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).

and their children, which are identified via the Swedish Medical Birth Registry. This sample constitutes approximately 98% of all births in Sweden during this period.

The Swedish Medical Birth Registry (Medicinska Födelseregistret) is provided by the National Board of Health and Welfare and contains detailed information on all pregnancies occurring since 1973 that led to a childbirth (at greater than 22 weeks of gestation) in Sweden. The information is provided by all Swedish prenatal care units, maternity clinics and neonatal care units, and contains extensive information regarding pregnancy, delivery, and health of the newborn child (including both stillborn and live births). These data include information on birth outcomes and child characteristics such as delivery mode, parity, multiplicity of births, fetal position, gender, gestational age, birth weight, length, head circumference, Apgar score[9] measured at 1, 5 and 10 minutes, malformations and severe maternal complications.[10]

The Medical Birth Registry also contains detailed information on maternal characteristics such as age, number of previous births, weight, height, chronic diseases, and tobacco consumption. It also contains information on prenatal conditions and treatments such as the use of fetal diagnosis service and pregnancy complications (diagnosis and procedures). Since 1995, the Medical Birth Registry collects information on fertility treatments including standard IVF, Intra cytoplasmic sperm injection (ICSI), surgical procedures and ovarian stimulation, when resulting in a successful pregnancy delivered after week 22. During the period of 1998-2007, 21,783 babies born after IVF are registered in the Medical Birth Registry. Out of those, 20% are twin births.

We combine the Medical Birth Registry with the National Patient Registry in order to obtain data on the number of nights spent in hospital by the mother and child (inpatient care), as well as the Cause of Death Registry to obtain information on mortality. Both registries are provided by the National Board of Health and Welfare. Finally, administrative data on income and educational attainment of mothers is obtained from the Social Insurance Agency (Försäkringskassan) and the Swedish Agency for Innovative Systems (LISA), provided by Statistics Sweden. This information includes income from gainful employ-

[9] Apgar score, 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.

[10] Severe maternal complications include postpartum hemorrhage (severe blood loss) and maternal sepsis (infection). 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).

ment, parental benefits and sickness benefits[11] as well as the highest level of education of each woman.[12]

3.2 Main outcome variables, data limitation and multiple hypothesis testing

In analyzing the effect of the reform we focus on mother and child health outcomes, and maternal labor market outcomes. As discussed above, child health outcomes include measures frequently used in the economic literature on early-life human capital (for example Apgar, birth weight, length, head circumference and gender, infant mortality, and mortality under the age of 5) (Almond et al., 2017; Björkegren et al., 2016), as well as an additional set of rich measures available in Swedish registry data (malformation, breech presentation at birth, nights hospitalized during the first year of life and during years 1-4). For mothers, we examine a series of variables capturing health at the time of child birth, and potential results of child birth on subsequent health. These are the use of emergency C-section (a C-section after attempting vaginal delivery), maternal sepsis, postpartum hemorrhage, hypertension and nights hospitalized during the first year after delivery. Finally, we examine maternal labor market outcomes, namely income from gainful employment, parental benefits and sickness benefits in expressed in real terms using 1980 consumer price index and 100s SEK. We use an inverse hyperbolic sine transformation $\log(y_i + (y_i^2 + 1)^{1/2})$, given that a non-negligible portion of women have zero income in the years considered, and this measure can be interpreted in a similar way to a log transformation, while also being defined at zero.

The Medical Birth Registry contains all births delivered after the 22$^{\text{nd}}$ week of gestation (plus 0 days). As such, *unsuccessful* IVF treatments are not observed. For this analysis we focus on births that are the product of IVF procedures including standard IVF and IVF with ICSI. However, based on official usage figures, the Medical Birth Registry only contains approximately 70-90% of all IVF births occurring during this period, misclassifying 10-30% of all IVF births as non-IVF births. This means that the control group consisting

[11]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.

[12]This is 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.

of non-IVF births are “contaminated” by a small number of IVF births. We return to this point in section 5.3.2, and document that even under conservative assumptions it is likely to cause only a very small attenuation of estimated reform impacts.

When examining the impact of the SET reform on child and maternal outcomes we are interested in multiple outcomes to capture child or maternal well-being. 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 labor outcomes separately. By doing so we decrease the number of hypotheses tested to a single outcome for each class of outcome variables. These indices are constructed as per Anderson (2008) by first ensuring that variables are consistently measured so that more positive values imply a positive change,[13] and 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 so that those contributing the most linearly independent information receive a higher weight in the index.

Secondly, 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 considerably more demanding Bonferroni (1935) corrected p-values, which controls for the Family Wise Error Rate, and thus sets the size of each test to avoid falsely rejecting *any* hypothesis.

3.3 Sample and summary statistics

We consider all twin and singleton births conceived between 1998 and 2007. We remove the small proportion of triplet and higher order births (516 births in all) given that these are a particularly extreme and uncommon outcome. This period consists of 60 months before and after the definition of the new SET guidelines relating to embryo transfer procedures. During this period there are 21,783 births following IVF recorded in the Medical Birth

[13]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.

Registry and 916,110 non-IVF births.

Table 1 displays summary statistics for maternal and child characteristics for IVF (columns 1-3) and non-IVF births (columns 4-6). We report t-tests for the equality of means of each variable between IVF and non-IVF births in column 7 and p-values in column 8. As expected, the rate of twin births is significantly higher among IVF births: approximately 20% of IVF births result in twins compared to 2.5% among non-IVF births. Based on observable characteristics and outcomes, women conceiving using IVF are different to women with unassisted conception along multiple dimensions. First, women conceiving with IVF are much more likely to suffer from pregnancy and birth complications such as hypertension, maternal diabetes, postpartum hemorrhage, maternal sepsis and emergency C-sections compared to non-IVF mothers. For example, mothers conceiving with IVF had double the risk of postpartum hemorrhage (12%) and emergency C-section (16%) compared to non-IVF mothers. IVF mothers are also more likely to be hospitalized the year after giving birth. Second, women conceiving with IVF are somewhat taller and have a slightly higher weight. Moreover, IVF mothers are on average older (age 33) than non-IVF mothers (age 30). The age distribution of IVF and non-IVF mothers is presented in Figure 7. IVF mothers have higher education attainment based on the categorical measure available (4.7 compared to 4.5) and higher labor income (average annual income before birth is 65.4 TSEK compared to 42.5 TSEK).[14] However, women conceiving with IVF receive more sickness benefits before giving birth, of 1.9 TSEK compared to 1.4 TSEK and are more likely to suffer from diseases like ulcerative colitis. Moreover, behavioral differences are found between women conceiving with IVF and without IVF: IVF mothers are more than 50% less likely to smoke cigarettes during the first (4.2% versus 9.6%) and third (2.4% versus 6.6%) trimester. These differences in means suggest that high SES mothers (higher education and non-smokers) select into IVF but also that women conceiving with IVF are more likely to suffer from chronic diseases and receive sickness benefits.

Differences in means between children born following IVF and non-IVF are displayed in Panel B, and show a similar pattern of poorer health outcomes. IVF children have lower Apgar scores (9.63 versus 9.73) as well as a higher likelihood of having Apgar scores below 7. Mortality rates (infant and under-5) is higher among children born after IVF with means of 4.5 and 5.6 compared to 2.7 and 3.4 for children with an unassisted conception. Other

[14]Expressed in real terms using 1980 consumer price index.

important health indicators show a similar pattern, including shorter gestation by nearly a week, lower birth weight by 300 grams, smaller head circumference and shorter length at birth. Malformations, breech presentation, neonatal hospitalization and hospitalization during ages 1-4 are higher among children born after IVF. No statistical difference is observed in the sex ratio.[15]

Previous studies suggest that twin births are a major contributor to the observed differences in outcomes between children born following IVF and those following unassisted conceptions (Kalra and Barnhart, 2011). However, singletons born after IVF have also been shown to exhibit poorer health outcomes compared to non-IVF births (Pinborg et al., 2013). Similarly, as documented in Table 2, we observe significant differences between singletons born after IVF and those not following IVF (in the pre-reform period). These differences are however smaller. Singletons born after IVF weigh 100 grams less than non-IVF births and exhibit alleviated risk of mortality and hospitalization. This provides suggestive evidence that a significant part of the differences in poorer health outcomes between IVF and non-IVF births is due to the higher prevalence of multiple births.

3.4 Descriptive statistics of the SET reform

Trends in the rates of twin births among IVF and non-IVF conceptions are displayed in Figure 3. A clear drop in rates of twin birth from 30% to 13% is seen among IVF births in line with the SET reform in 2003, while no similar change in rates of twinning are observed for non-IVF births during this period. The improvements in child health for children born after IVF is seen across multiple outcomes and displayed in Figures 4 and 5. These graphs show a clear pattern of improved health for children born after IVF compared to non-IVF children when considering birth weight, gestation, length, head circumference and Apgar score as well as reductions in mortality and the probability of hospitalization within one year of birth. Trends in maternal health and labor market outcomes are presented in Figure 6, and show a similar pattern of improved outcomes for the likelihood of emergency C-section as well as somewhat decreased sickness benefits and higher labor income.

A before and after comparison of differences in means in outcome variables for IVF births during 1998-2007 is presented in Table A2 and confirms the observed trends. Means are reported, with standard deviations below each mean, and an associated *t*-test in column

[15] Male fetuses are less resilient to more demanding conditions in utero (Almond and Mazumder, 2011).

3. As in the graphical evidence, significant differences in means around the reform are found for multiple child outcomes including birth weight, gestational age, length and head circumference as well as a lower probability of mortality. In terms of maternal outcomes, higher labor income and lower sickness benefits are observed following the SET reform, and health improvements are observed, for example a lower prevalence of emergency C-sections.

4 Empirical strategy

We estimate the impact of the SET reform using the following difference-in-differences (DiD) specification:

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

This exploits both variation in IVF usage and reform timing, comparing outcomes for IVF and non-IVF births prior to and posterior to the January 1, 2003 policy change. 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). The parameter of interest is β_1, capturing the change in outcomes for IVF births relative to non-IVF births after relative to before the reform was implemented. While obstetric outcomes among IVF births are expected to be better post-SET, IVF children will nevertheless tend to have poorer obstetric outcomes than children born following an unassisted conception (Sazonova et al., 2011). Our estimates will allow us to capture not only the SET-led improvements in IVF outcomes but also the extent to which SET led to a convergence of outcomes from IVF with outcomes from non-IVF. Here $PostSET$ is a binary variable based on estimated date of *conception*: all births estimated to have been conceived after January 1 2003 are assigned as $PostSET = 1$.[16] Rather than include the uninteracted $PostSET$ term in the regression, we include a series of year fixed effects π_t to flexibly control for any time varying unobservables that may have evolved in a manner similar to the reform. County-specific fixed effects α_c capture time-invariant geographical variation in the outcomes.

In some specifications, we additionally include maternal and birth characteristics, de-

[16]Conception date is computed by subtracting the gestational days from the date of birth analogous to Currie and Schwandt (2013). Although date of birth is not available in our data set, we use the discharge date for the maternity unit.

noted $\boldsymbol{X}$. These include age and pregnancy order fixed effects, maternal height and weight before pregnancy, nationality (a binary variable for having been born in Sweden or not), whether the mother smoked during the first trimester of pregnancy, and the mother's educational level, sickness benefits and labor income averaged over the 3 years prior to birth. The idiosyncratic error term is denoted by ε, and is clustered on the mother. We estimate equation 1 using OLS.

The identifying assumption 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 "treatment" indicator (IVF) with a binary variable for each year prior and posterior to the reform date. The specification we estimate is:

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

where $\ell = \{-4, -3, -2, 0, \ldots, 4\}$ and the year before the SET reform, 2002, is omitted as a base category. Equation 2 is similar to equation 1 except that instead of defining differences around a single post-SET binary variable we allow the difference between IVF and non-IVF births to vary year on year. If IVF and non-IVF outcomes exhibit differential pre-trends then this will be evident in a test of the lagged coefficients.

We estimate the reduced form impact of the SET reform, that is, the average treatment effect among all IVF births (the intent to treat estimate). While the substantive change brought about by SET is a reduction in twinning and this is the main mechanism for impacts on outcomes, maternal selection into SET may also play a role. We expect positive selection into SET since women perceived to have a low risk of twinning (older women and/or multiple previous failed IVF cycles) were allowed to elect for DET following the SET reform. Positive selection is a concern as it will tend to lead to overestimation of the improvements in child and maternal outcomes. To account for selection into SET we estimate a specification that conditions upon mother fixed effects. However, only approximately 50% of all IVF mothers in the sample have more than one birth. We therefore show results on the restricted sample of women with and without mother fixed effects. This way we can isolate changes in the estimates arising from selection of a sample of women with at least two births from changes in the estimates associated with selection into SET. As one check on the twinning channel, we estimate a regression of the outcomes

Y_{it} on whether the birth is a twin birth, instrumenting the indicator for a twin birth with an indicator for whether the birth occurred post-SET. This provides a local average treatment effect (LATE) for SET compliers.

We will subject the estimates to a number of robustness checks. We will discuss the fact that SET was mandated two years earlier in one county and re-estimate the model excluding this county. We also re-estimate it 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. We will investigate changes in the composition of mothers selecting into IVF treatment after SET, although this is accounted for by the main effect of IVF in our specification.[17] We will investigate heterogeneity in impacts of the reform by mother characteristics. Of particular interest, we will show all results for all women and then again for first-time mothers (44% of the sample).

5 Results

5.1 Twin births

Table 3 presents the impact of the SET reform on the likelihood of a twin as opposed to a singleton birth. For the full sample (columns 1-2) we estimate a reduction in the share of twins among IVF births of 16.8 to 17.3 percentage points (pp), depending on whether we do or do not control for the mother's characteristics. Estimates for first-time mothers (columns 3-4) are very similar at 17.7 pp, which is not sensitive to controls for woman characteristics. Using the twin rate among IVF births in the pre-treatment period of approximately 27 %, our estimates indicate that the SET reform narrowed the gap in twinning between IVF and non-IVF births by about 63%. To account for omitted trends that are specific to IVF outcomes, 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 A3). The general pattern of the results is maintained but the reduction in twinning is closer to 13 pp rather than about 17 pp.

We investigate impacts of SET on twinning by sub-groups identified by birth order, the mother's age at treatment, her education and her BMI; see Table 4. We see a statistically significant reduction in the share of twins among IVF births in every sub-group except

[17] As explained in the Data section, our data do not allow us to investigate differences in characteristics of women who post-SET end up electing for SET vs DET, but we discussed above how we account for selection into SET among IVF users.

for women 40 years and older. As these women have a lower probability of twinning (see Panel B of the table: the probability is 14% compared with about 27% on average), they were probably exempt from SET. Estimates by birth order show that the impact of SET is smaller for births of order 3 or higher, estimates by age show that the impact of SET on twinning is hump-shaped in age, being smaller for women under 25 and women 40 or older. There are no significant differences by the woman's education, but impacts are smaller for women with low BMI relative to other women.

5.2 Child and maternal outcomes

Table 5 presents estimates of the impact of the SET reform on child health, maternal health and maternal labor market outcomes for the full sample (columns 1-3) and first-time mothers (columns 4-6). We identify a significant improvement in the index of health of 0.189 standard deviations (SD), which is similar for first-time mothers and all mothers. This implies that the SET reform reduced the health gap between IVF and non-IVF children of -0.355 SD by 53%. The impact of SET on the maternal health index falls just short of significance in the sample of all mothers but it is statistically significant for first-time mothers, for whom health improves by 0.056 SD. This 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 other mothers). Maternal labor market outcomes within 3 years of birth improve by 0.106 SD for the full sample and by 0.156 for first-time mothers. Consistent with extensive margin fertility (the first birth) having larger impacts on labor market outcomes, the IVF/non-IVF gap in labor market outcomes 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.

We present 2SLS estimates using the passage of the SET reform to instrument the likelihood of giving birth to a singleton in Table 6. We thus examine the (SET mediated) impact of having a singleton birth rather than a twin birth on child health, maternal health and maternal labor market outcomes. This is the local average treatment effect on compliers: women who had a singleton birth because they had IVF treatments after the policy change, but who would have had twins if SET were not the default policy. In columns 1-2, the first stage results are presented, which show the strong reduction in twinning for the full sample and first-time mothers, with F-statistics far exceeding typical weak instrument thresholds. For the full sample, the 2SLS estimates suggest a strong and significant impact of having a singleton child on the child's own health (1.1 SD of the index) as well as on maternal labor market outcomes (0.63 SD of the index). For first-time mothers, the 2SLS estimates suggest that having a singleton birth compared to twin births causes a strong positive impact on child health (1.02 SD), maternal health (0.31 SD) and labor outcomes (0.87 SD).

Examining each child health outcome separately in Table 7, for the full sample (Panel A) and first-time mothers (Panel B), we find that the improvement in child health is driven by multiple factors. In particular, the reform led to an increase in the average absolute Apgar score (column 1), a lower probability of having an Apgar score below 7 (column 2), increased birth weight (column 3), increased length of the baby at birth (column 4), a larger head circumference (column 5), longer gestation (column 6), declines in infant and under 5 mortality (columns 8 and 9), a lower likelihood of breech presentations (column 10) and a lower probability of hospitalization during first year of life (column 11). We find no effect on the child's gender, rates of malformation or hospitalization during ages 1-4. The magnitudes of these effects are large. For example, average birth weight increases by 175 grams, closing the gap between IVF and non-IVF babies by 57%. Similarly the gestational age increased by more than half a week following the reform, closing the gap by 52%. Changes in birth length and head circumference also reduced the IVF–non-IVF differential by 50%. These findings are of interest given the well-documented causal relationships between birth weight, gestational age, length and head circumference with later life outcomes (Bharadwaj et al., 2013; Björkegren et al., 2016). To account for multiple hypotheses when examining individual child-health components, we correct the p-values with a false discovery rate as well as Bonferroni correction. Even when correcting p-values with the Bonferroni correction—a particularly demanding test—highly significant effects on Apgar score, birth weight, length, gestation, head circumference, hospitalization and breech presentation remain. Similar results are found for first-time mothers but with slightly larger effects.

We estimate the impact of the SET reform at particular points in the distribution of birth weight and gestational length that are commonly used in the literature, presented in Table 8. Results are presented for the full sample in Panel A and first-time mothers in Panel B. We focus on these weights/dates given their importance in the targeting of medical resources based on (arbitrary but commonly used) treatment cut-offs (Almond et al., 2010; Bharadwaj et al., 2013). For the full sample, the impact on the likelihood of being born with a weight below 1500 grams (very low birth weight), is large and negative, at 1.2 pp, and for a weight below 2500 grams (low birth weight) is a 6.8 pp reduction (column 2). This corresponds to a decrease of 60% when compared to the rate of low birth weight babies born via IVF before the reform. Similarly, the probability of preterm delivery before week 28 (column 3), 32 (column 4) and 37 (column 5) decreases by 0.5, 1.3 and 8.3 pp, closing the gap by around 50% in each case. Importantly, these results demonstrate that average impacts on birth weight and gestation are not driven only by changes on the upper quintiles of outcome distributions. Very similar results are found for first-time mothers (Panel B).

Turning to maternal outcomes, the results for each separate health outcome from the maternal health index are presented in Table 9. The results are presented for the full sample (Panel A) and first-time mothers (Panel B). These results demonstrate no significant impact on complications such as postpartum hemorrhage (column 2), maternal sepsis (column 3), post birth hospitalization (column 4), or hypertension (column 5) for either the full sample (Panel A) or for first-time mothers (Panel B). Unsurprisingly, we do observe a significant negative impact on the likelihood of engaging in an emergency C-section at birth (column 1), and this impact closes the gap between IVF and non-IVF births by 42% for the full sample and 60% for the first-time mothers. The negative impact on emergency C-section remains highly significant when correcting for multiple hypothesis testing using a FDR and Bonferrroni correction.

We document the impact on each labor market outcome separately in Table 10, again presented separately for the full sample (Panel A) and first-time mothers (Panel B). We have transformed each income variable using hyperbolic sine transformation, so each coefficient can be interpreted as a percentage change. Here we observe that income from parental benefits decreases by 3.4%, sickness benefits decline by 43.6% and labor income increases by 8.5%. These findings suggest that in the three years following the birth of a child, IVF mothers have significantly higher labor earnings and lower usage of benefits and transfers. Correcting for multiple hypotheses tests renders no longer a significant effect on parental benefits and labor income when using the conservative Bonferroni correction (but not when referring to q-sharpened p-values). For first-time mothers, a similarly strong decrease in sickness benefits is found by 46.0%, but with no significant impact on labor income or parental income within 3 years after giving birth.

5.3 Identifying assumptions

5.3.1 Parallel trends and event studies

As is standard in difference-in-differences analyzes, correctly identifying the impact of the reform requires a parallel trends assumption. In this case, we must assume that trends in outcomes among IVF and non-IVF women would have evolved similarly over time in the absence of the reform. While this assumption cannot be tested directly, we can partially test its plausibility using event studies to examine the evolution of outcomes in the IVF and non-IVF groups in the pre-reform period.

In Figure 4, we present trends in birth weight, gestational age, length, head circumference and Apgar score, which show a clear improvement following the SET reform (the reform date is indicated by the red vertical line). Trends in mortality and hospitalization exhibit larger variation but also show an improvement following the 2003 reform. No apparent decrease is seen in malformations and hospitalization during ages 1-4. While some outcomes exhibit larger variation, overall, these graphs suggest approximately parallel trends in the outcome variables prior to the reform by simple visual inspection. While trends in maternal labor market outcomes i.e. income, parental, and sickness benefits appear to be parallel in the pretreatment period depicted in Figure 6, trends in maternal health exhibit large variation making it hard to assess the presence of common trends in the pre-treatment period presented in Figure 6.

To examine if the assumption of parallel trends is satisfied, we formally test this by IVF and reform lags and leads, as per equation 2. By allowing for a more flexible model we can infer trends in the pre-treatment period as well as whether the effect is persistent in the post-reform period. The event studies are presented in Figure 8, and confirm previous findings of a sharp and persistent decline in twin births (Figure 8a). Similarly, we find a strong increase in both child health (Figure 8b) and maternal labor market outcomes (Figure 8d) but no impact on maternal health (Figure 8c). Event studies for each component in the indices are presented in Figures A2, A3, and A4. In terms of pre-trends, twin births exhibit two coefficients in the pre-reform period which are significantly different from zero, suggesting some fluctuations in the pre-reform period. The maternal labor market index is not significantly different from zero in the pre-reform period suggesting that we cannot reject the absence of parallel trends in labor market outcomes prior to the SET reform. The event study of the child health index also suggests parallel trends in the pre-reform period.

We further examine if the results are robust to the inclusion of linear time trends, presented in Table A4, using both split trends (allowing for different slopes across the time of the reform) for IVF and non-IVF births (Panel A) and global trends for IVF and non-IVF births (Panel B). For the full sample, the results are consistent with the baseline results but with somewhat smaller coefficients for child health index and maternal labor index (of 0.137 and 0.061 respectively). The exception is the maternal health index, which suggests a significant positive effect of the magnitude 0.106 SD. For first-time mothers, the results are similar to the baseline, but with somewhat smaller magnitude for child health and labor outcomes (0.097 and 0.105 SD) and larger magnitude for maternal health. In Panel B, the results when including IVF specific linear trends are presented for the full sample (columns 1-3) and show a similar result to the baseline result for child health but with a smaller and less statistically significant impact on the maternal labor market index and with a positive impact on the maternal health index. For first-time mothers, results are similar to the baseline but with a stronger impact on the maternal health index. These results show that the estimated impact of SET is overall robust to the inclusion of trends but with the exception for maternal health outcomes, which indicates a positive significant impact when controlling for trends.

5.3.2 Compositional changes in IVF mothers and maternal selection to SET

The proportion of IVF births has increased since the 1990s, tracking changes in technologies, costs, and availability of IVF (see Figure 1b). It is likely that the composition of mothers using IVF also changed throughout this time. In order for this to invalidate our identification strategy, the composition of mothers must evolve differently for IVF and non IVF users around the date of the SET reform. To further explore this, we examine possible compositional changes using observable maternal characteristics including age, height, weight, education, labor and sickness benefits prior birth, nationality, smoking, asthma, epilepsy and ulcerative colitis. That is, we perform a balancing test of covariates across the time of the reform by running regressions using maternal characteristics as outcome variables and the SET reform as the explanatory variable. The results are presented in Table 11, and show a significant impact on three outcomes out of eleven. These variables are: age (column 1), education (column 4), and smoking (column 8). While the magnitude of the coefficient on age is rather small, the magnitude of the estimates for education calls for closer consideration. It is hard to assess how a potential change in "quality" in mothers have evolved over time. These estimates suggest that mothers are both slightly older and somewhat more educated but also slightly more likely to be smokers.

We investigate the influence of this on our estimates of the impacts of SET using three complementary approaches. First, we re-estimate the baseline results including linear trends in maternal characteristics and IVF-status in each model. In Table 12, the estimates are presented for the full sample (columns 1-3) and first-time mothers separately (columns 4-6). For the full sample, the results show a positive and significant impact on child and labor market outcomes, as in the baseline specification. The estimate for child health (0.149 SD, column 2) is similar to the baseline estimate but for the maternal labor market index the estimated coefficient is smaller (0.059 SD, column 3). In contrast to the baseline results, a positive and significant effect is found for maternal health (column 2) of approximately 0.112 SD, which suggests that maternal health may have improved by the reform. For first-time mothers, the result is similar to the baseline but with a somewhat smaller magnitude in child health (0.122 SD, column 4) and somewhat larger impact on maternal health (0.105 SD column 6).

Since we can only control for trends in observable characteristics, we also examine estimates based on mother fixed effects. These will control for all unobservable time-invariant characteristics of mothers. To implement this model we need to restrict the sample to women with at least two pregnancies.[18] Approximately 50% of all IVF mothers have more than one pregnancy, and as such, the use of mother FEs excludes half of the sample. Since mothers with two pregnancies may not be representative of all mothers, we examine characteristics of IVF mothers with one pregnancy versus two pregnancies, and these are presented in Table A5. 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

[18]We have shown that the SET reform led to a highly significant decline in twin births among a broad category of women e.g. all education levels, BMI classifications, parity and ages except for women older than 39. This suggests that compliance with the SET reform is higher among younger women. Provided that child health is negatively correlated with rising maternal age, this selection could bias our estimates upwards. In addition, there is likely to be maternal selection based on unobservable characteristics, where mothers with multiple previously failed IVF cycles are less likely to comply with SET, and are at a greater risk of adverse birth outcomes. That is, maternal selection into SET could potentially lead to biased estimates, and in particular, may cause *upwardly* biased estimates of the improvements in child health and maternal labor market outcomes.

unobserved mother-level heterogeneity.

The results are presented in Table 13 and show the impact of the SET reform on twin rates (column 1), child health (column 2), maternal health (column 3) and labor market outcomes (column 4). In Panel A, the impact of SET on the selected sample of IVF mothers excluding mother fixed effects are displayed, and show that the coefficients are similar to the baseline results. In Panel B, results are presented including mother fixed effects. Again, similar results to the baseline results are observed but with somewhat lower precision when including mother fixed effects. We continue to observe large and significant effects of the SET reform even when only examining *within*-mother variation in IVF laws. The impact of the SET reform is estimated to increase child health and labor market attachment of women following birth.

To assess the magnitude of a potential selection bias, we postulate that selection based on observable explanatory variables provide information on selection on unobservables as suggested by Altonji et al. (2005). We consider the magnitude of omitted variable bias needed to eliminate the impact of the SET reform, by computing the ratio of how large the covariance between unobservables and the SET reform and the covariance between observables and the SET reform must be to explain away the impact of SET.[19] The more the inclusion of controls is affecting the coefficient indicating the treatment effect, the larger the potential bias is and vice versa. A large ratio indicates that it is less likely that omitted variables would explain away the impact of the reform. Table 14 presents the results without including maternal controls. The coefficients are very similar with and without controls, which is indicated by the ratio between the two regression models presented in the bottom row. This simple exercise suggests that, given the limited selection on observables (showed by the ratio) we may assume that selection on unobservables is equally limited.

5.4 Mechanisms

We examine potential mechanisms through which maternal labor market outcomes are affected by examining the importance of different components of the reform. We consider the negative fertility shock as a direct effect of the reform and child and maternal health as indirect effects. As a mechanism test we include child and maternal health outcomes,

[19]This can be computed by using the estimates from OLS regressions both with and without controls: $\frac{\alpha_{controls}}{\alpha_{nocontrols}-\alpha_{controls}}$. For a more detailed discussion see Bellows and Miguel (2009).

which themselves are outcomes of the reform and therefore "bad controls" (Angrist and Pischke, 2009). However, they will provide us with information on how much the reform is affecting maternal labor outcomes when improvements in child and maternal health are controlled for. To do this we adopt the conditional decomposition proposed by Gelbach (2016). We are interested in how much the estimated effect of the SET reform is affected by 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

$$\begin{aligned} &\beta^{uconditional}_{labor} - \beta^{conditional}_{labor} \\ &\qquad = \Gamma^{childhealth}_{labor}\beta^{childhealth}_{labor} + \Gamma^{maternalhealth}_{labor}\beta^{maternalhealth}_{labor} \\ &\qquad\qquad = \delta^{childhealth}_{labor} + \delta^{maternalhealth}_{labor} = \delta_{labor} \end{aligned} \tag{3}$$

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.

In Table 15, the potential mechanisms of child and maternal health are included in the baseline specification, Equation 1. Table 15 presents the impact of the SET reform on child health (column 1), maternal health (column 2) and maternal labor index including the potential mechanisms of child and maternal health (column 3). To see how each of the components are affecting the maternal labor index, the Gelbach decomposition is presented in column 4. The decomposition shows that the impact of changes in child health owing to the reform explains only 0.005 of the improvements in the maternal labor market index, with an even smaller value of 0.002 owing to changes in maternal health. The total explained difference is 0.007 of the total 0.099 SD of improvements in the maternal labor market index following from the SET reform. This suggests that the drop in fertility is the main contributor to improvements in maternal labor outcomes.

5.5 Additional robustness and sensitivity

Additional robustness checks are presented in Table 16. First (Panel A), we remove births from the two years prior to the SET reform (2001-2002) in order to account for a potential gradual increase in SET, which may bias our estimates downwards because of a partially contaminated control group. The impact of SET on twin birth (column 1), child health index (column 2), maternal health index (column 3) and maternal labor index (column 4) shows a similar coefficient and effect size to the baseline results. 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 remain largely similar to the baseline results when removing this region.

In 2005, Sweden started to offer same-sex couples publicly funded access to fertility treatments including IVF. Previous literature suggests that same-sex couples exhibit a higher socioeconomic status (Ahmed et al., 2011a,b). Their children, however, exhibit somewhat worse birth outcomes in terms of lower birth weight, when compared with children born to heterosexual couples (Aldén et al., 2017). While the number of children born to lesbian parents during 1995-2010 is only 750, we further examine a potential impact of this legislative change. We restrict our sample to conceptions occurring during 1998-2004, and re-estimate our baseline model. The result is presented in Panel C in Table 16. The results are similar to the baseline results but with somewhat smaller estimates of -15 pp for twin births, 0.154 SD for child health and 0.067 SD for maternal labor index.

As discussed previously, the Medical Birth Registry correctly identifies approximately 70% of all IVF births based on reported usage in aggregate national IVF data. We may be concerned that approximately 30% of IVF births are mis-reported, and are incorrectly reported as non-IVF births, thus contaminating the control group. In practice, given that the size of the "treatment" group is much smaller than the size of the "control" group, even if the reform's impact was very-large, the 30% of mis-classified IVF births will be unlikely to impact averages in the control group in a 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% suggests that there are 9,356 IVF births incorrectly classified as non-IVF births. This is only slightly over 1% of the entire group of births assumed to be non-IVF births. We provide additional discussion, as well as a calculation of the (small) magnitude of any expected attenuation, in Appendix B.

6 Conclusion

The invention of IVF allowed radical changes in the fertility behavior of some women and families, providing the opportunity to postpone childbearing, as well as assisting involuntary childlessness. However, there are also well-documented immediate and long-run costs associated with IVF-usage. As well as the direct financial costs of procedures, IVF births have been documented to be considerably more likely to suffer from adverse health outcomes when compared to non-IVF births (Saldeen and Sundström, 2005). The adverse health outcomes following IVF are mainly attributed to the increased likelihood of multiple births. Twin births are widely documented as a major risk factor for mothers and children, for example, given the alleviated risk of preterm birth, low birth weight, fetal malformation and complicated delivery (Gelbaya et al., 2010). In particular, premature delivery is associated with higher mortality as well as long term adverse effects on neurological development (Gelbaya et al., 2010).

In this study we document the causal impact of a reform mandating single embryo transfer (SET) for IVF treatments on a broad set of child and maternal health and labor market outcomes. Using rich Swedish registry data for the time period 1998-2007, we find that the SET reform led to a sharp drop in rates of twin births: by over 60% for women under age 40. Reduced rates of twinning are observed for a broad category of women across birth order, education level and BMI classification. We find a highly significant and sizeable impact of the SET reform on child health as measured by Apgar score, gestational age, birth weight, length and head circumference. These findings are important given the well-known links between human capital at birth, and outcomes across the entire life-course including cognitive and non-cognitive ability, educational attainments, health and life expectancy. Moreover, our results suggest a decrease in complications of labor such as breech presentation at birth, emergency C-sections and reduced usage of neonatal hospitalization. Overall, our estimates suggest that the adoption of SET as the official IVF procedure reduced the health differential between IVF and non-IVF births by over 50%.

We also find that the adoption of SET resulted in a sizable and significant impact on maternal labor outcomes. The positive impact on labor market outcomes originate from a reduction in the usage of sickness benefits, and increased labor income in the period three years subsequent to birth. A significant impact on maternal health is found for first-time mothers, closing the gap in maternal health between IVF and non-IVF mothers by 36%.

The large magnitude of health benefits from SET is not limited to the children, mothers and families, but will also have a positive effect on the health care system and social safety net. Improvements in health at birth and during gestation will have follow-on effects, reducing demand for prenatal, obstetric and neonatal care. The SET reform is likely to reduce the long-term costs associated with IVF procedures, but at the immediate cost of less choice for women and couples seeking IVF. However, given that the delivery rate was unchanged despite the shift from DET to SET, this suggests there is no fertility-cost to the increased health of the child at birth. One reason why multiple embryo transfer is still common in many industrialized countries is likely due to financial incentives, jeopardizing the health of children and mothers (Karlström and Bergh, 2007). IVF has increased rapidly since the 1980s and is now a key feature of the reproductive landscape and is likely to increase further in the near future. Any improvements in individual and aggregate health due to the adoption of SET as a default IVF option will be magnified as rates of IVF use increase.

References

Aaronson, D., R. H. Dehejia, A. Jordan, C. Pop-Eleches, C. Samii, and K. Schulze (2017): "The Effect of Fertility on Mothers' Labor Supply over the Last Two Centuries," .

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.

Abramowitz, J. (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. Currie, and V. Duque (2017): "Childhood Circumstances and Adult Outcomes: Act II," Tech. rep., National Bureau of Economic Research.

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.

Almond, D. and B. Mazumder (2011): "Health capital and the prenatal environment: the effect of Ramadan observance during pregnancy," *American Economic Journal: Applied Economics*, 56–85.

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.*

Angrist, J. D. and J.-S. Pischke (2009): *Mostly harmless econometrics: an empiricist's companion*, Princeton University Press.

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., 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. and D. Clarke (2016): "The twin instrument," .

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.*

Björkegren, E., A. Bütikofer, G. Conti, M. Palme, and K. Salvanes (2016): "Consequences of Health at Birth," .

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.

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

Browning, M. (1992): "Children and household economic behavior," *Journal of Economic Literature*, 30, 1434–1475.

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.

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.

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.

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

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.

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.

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.

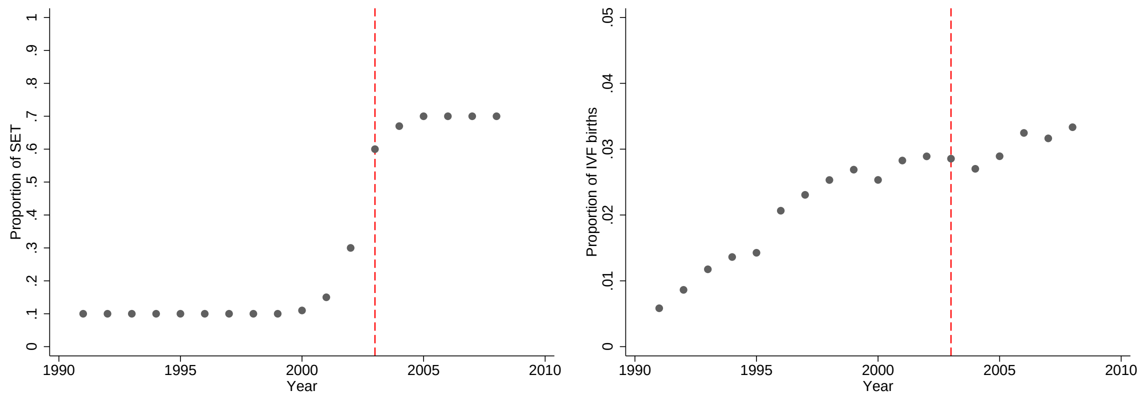
Vårdguiden (2017): "IVF, provrörsbefruktning," *1177 Vårdguiden.*

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.

7 Figures and tables

Figure 1: Trends in SET and proportion of IVF births

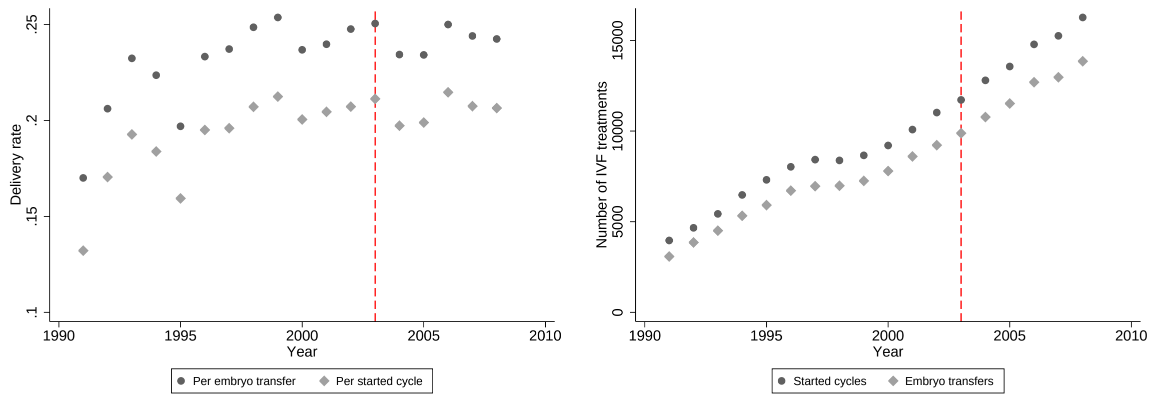
(a) Single embryo transfer **(b)** Proportion of IVF births



Annual trends in SET and proportion of IVF births are based on aggregated 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.

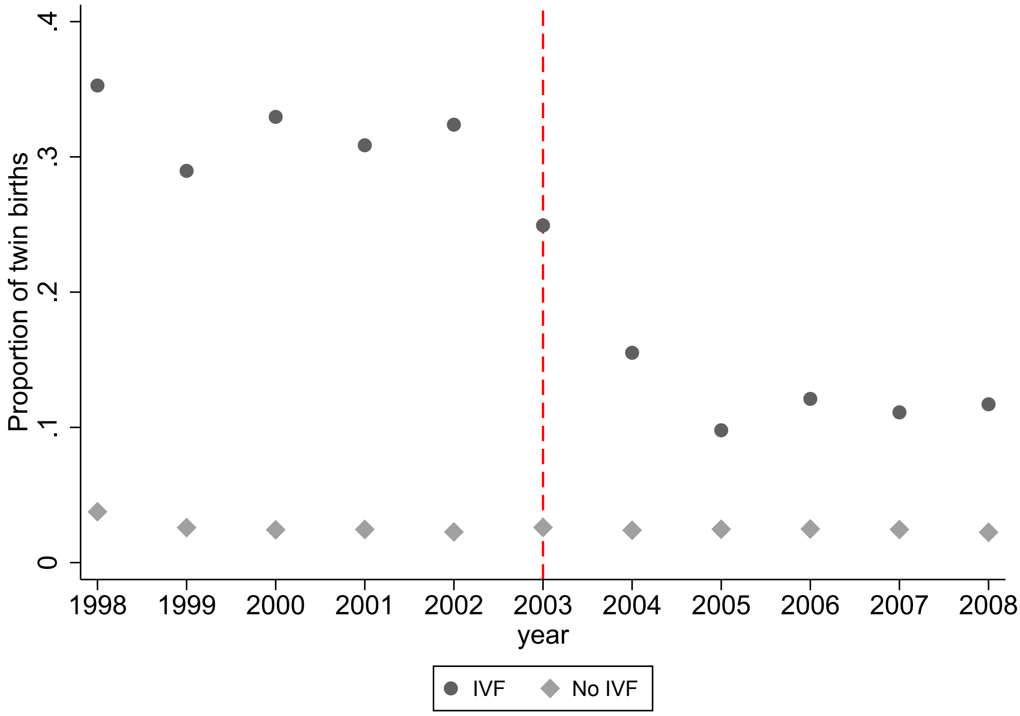
Figure 2: Trends in delivery rate and IVF treatments

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



Annual trends in deliveries per transfer/cycle and the number of IVF treatments are based on aggregated 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.

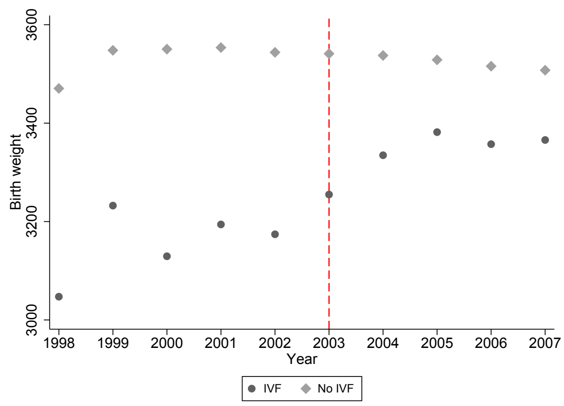
Figure 3: Trends in twin rates



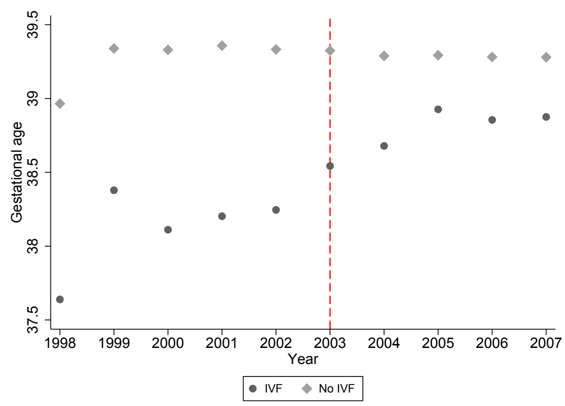
Annual trends in twin births with and without IVF conception, using data obtained from the Swedish Medical Birth Registry, are presented in Figure 3. The red vertical line indicates the year of the SET reform.

Figure 4: Child health outcomes

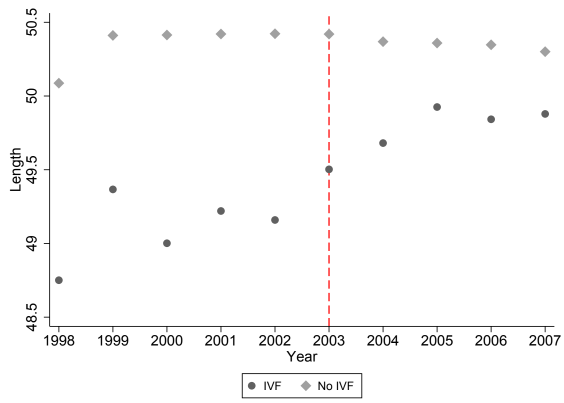
(a) Birth weight



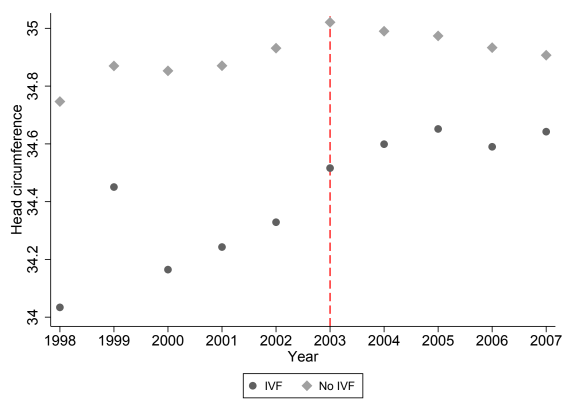
(b) Gestational age



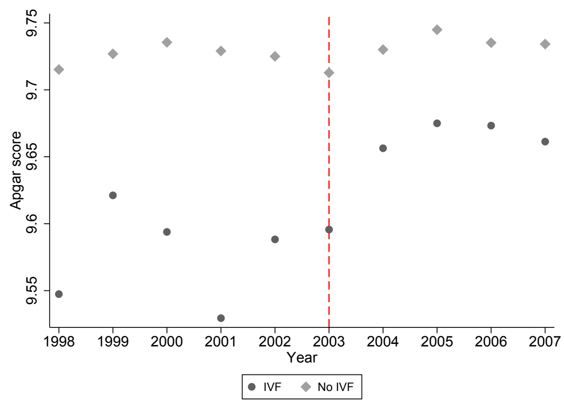
(c) Length



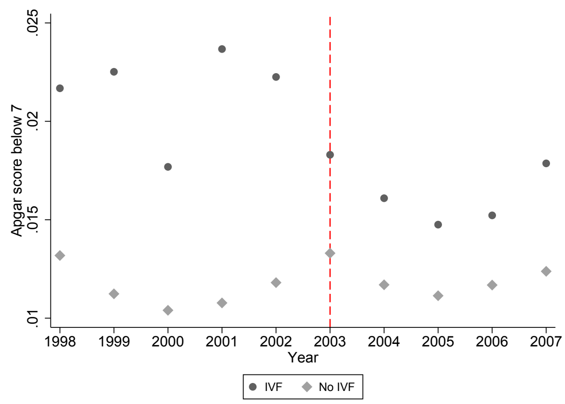
(d) Head circumfernce



(e) Apgar score



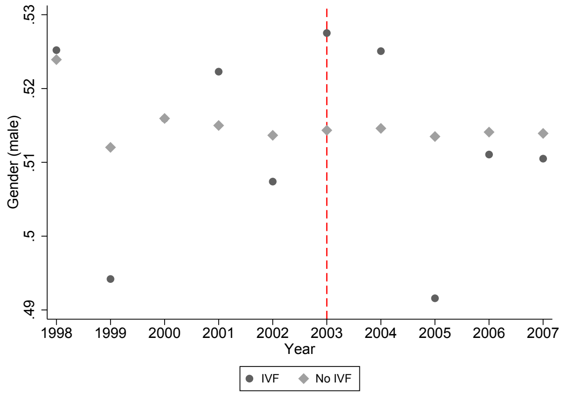
(f) Apgar score below 7



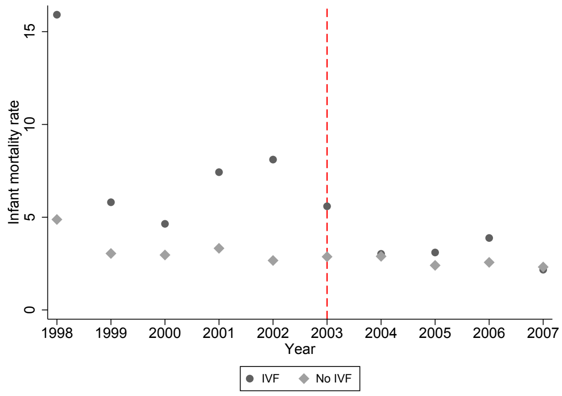
Annual trends in child health outcomes with and without IVF conception, using data obtained from the Swedish Medical Birth Registry and Patient Registry, are presented in Figures 4a to 4f. The red vertical line indicates the year of the SET reform.

Figure 5: Child health outcomes

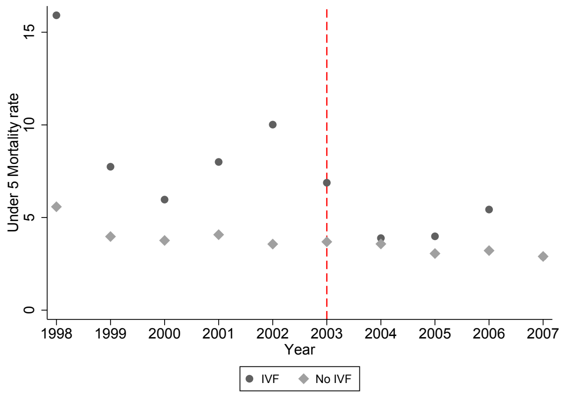
(a) Gender (male)



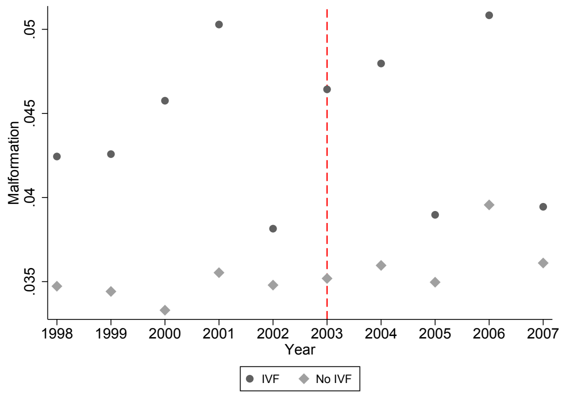
(b) Infant mortality



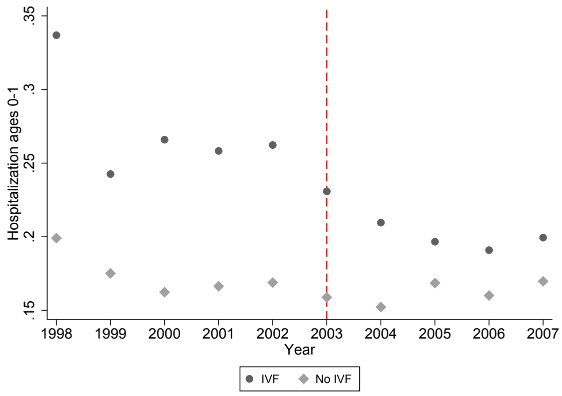
(c) Under 5 mortality



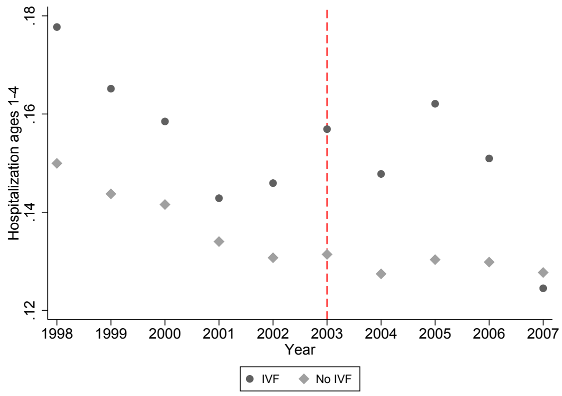
(d) Malformation



(e) Hospitalization ages 0-1

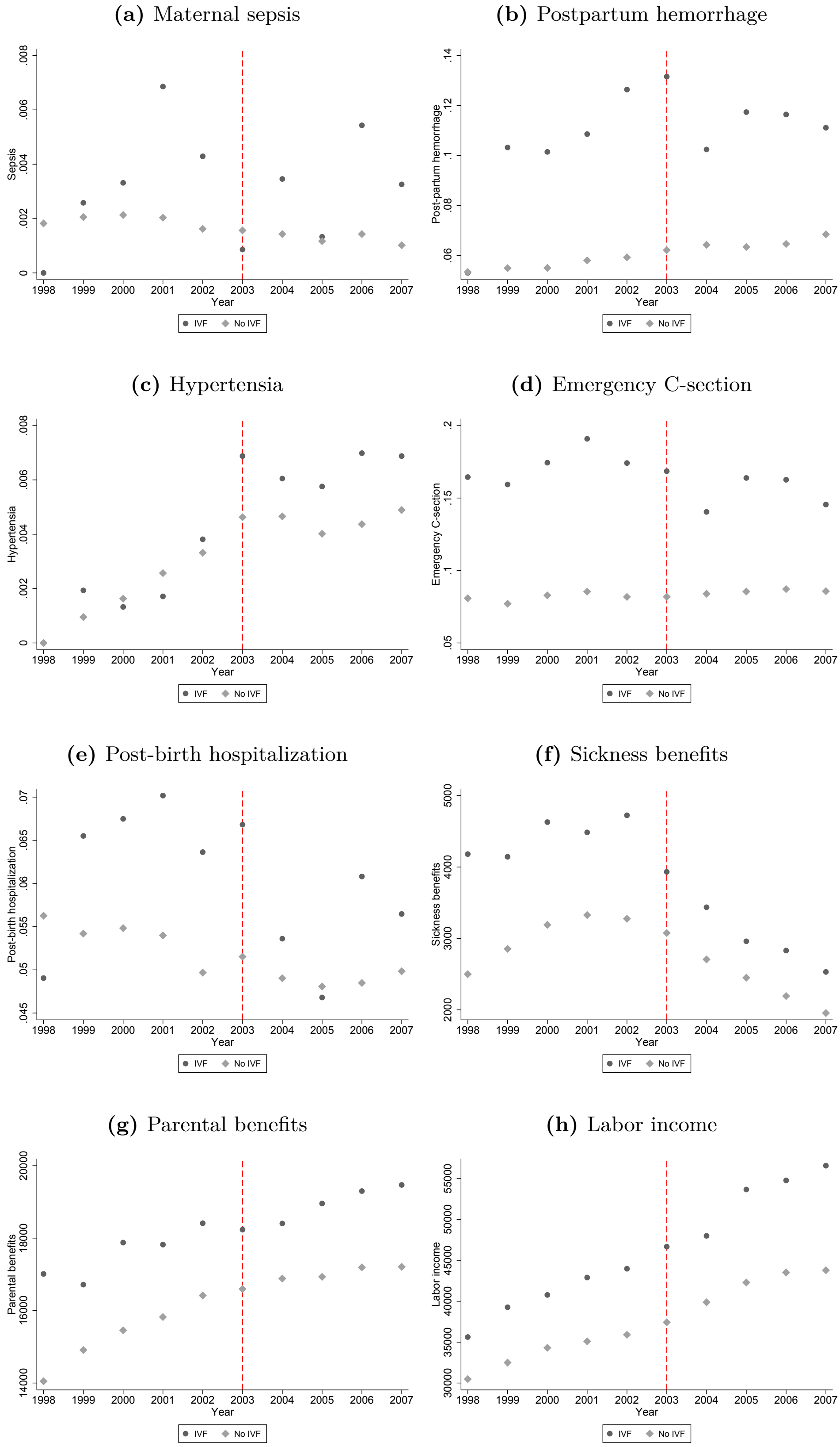


(f) Hospitalization ages 1-4



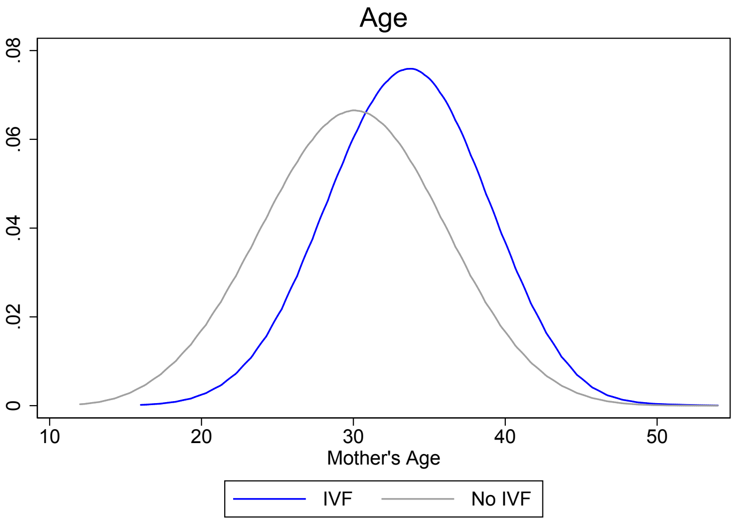
Annual trends in child health outcomes with and without IVF conception, using data obtained from the Swedish Medical Birth Registry and Patient Registry, are presented in Figures 5a to 5f. The red vertical line indicates the year of the SET reform.

Figure 6: Maternal health and labor outcomes



Annual trends in maternal health outcomes with and without IVF conception, using data obtained from the Swedish Medical Birth Registry and Patient Registry, and Longitudinal integration database for health insurance and labor market studies (LISA) are presented in Figures 6a to 6h. The red vertical line indicates the year of the SET reform.

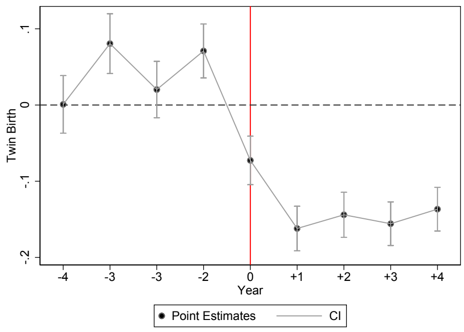
Figure 7: Age distribution among IVF and non IVF-mothers



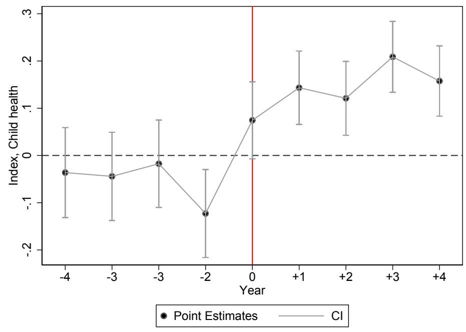
The data are obtained from the Swedish Medical Birth Registry for the time period 1998-2007. Figure 7 displays the age distribution among IVF and non-IVF mothers.

Figure 8: Event studies: main results

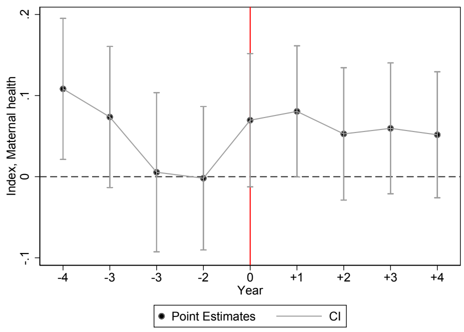
(a) Twin birth, full sample



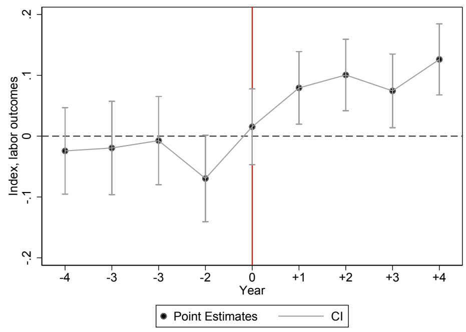
(b) Child health index, full sample



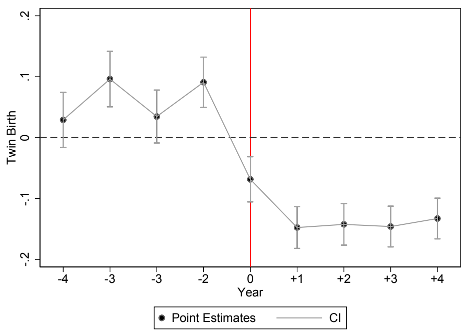
(c) Maternal health index, full sample



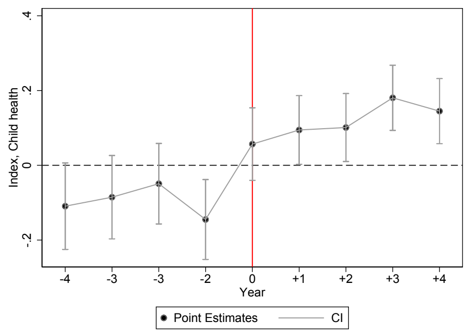
(d) Labor market index, full sample



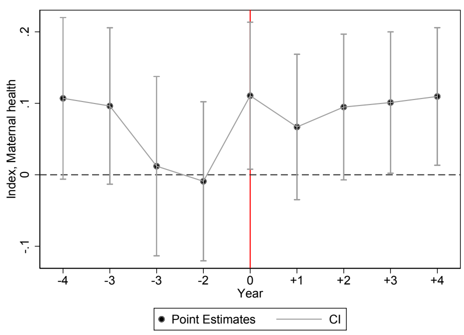
(e) Twin birth, first-time mothers



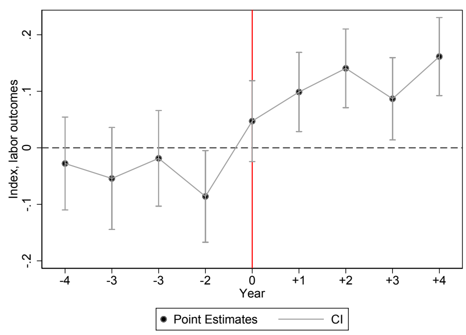
(f) Child health index, first-time mothers



(g) Maternal health index, first-time mothers



(h) Labor market index, first-time mothers



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 figure presents coefficients of interactions between each year and IVF births. The red-vertical line represents the year of the SET reform using the previous year as the omitted category. A full set of maternal controls and fixed effects are included in all regressions (as described in Table 3). Standard errors are clustered on the mother.

Table 1: Summary statistics

	IVF			*Non-IVF*			T-test	
	Mean	SD	N	Mean	SD	N	T-test	P-values
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Maternal outcomes and characteristics								
Twin birth	0.199	0.400	21783	0.025	0.155	916110	-154.479	0.000
Emergency C-section	0.162	0.368	21782	0.084	0.277	916102	-40.785	0.000
Maternal sepsis	0.003	0.056	21783	0.002	0.039	916110	-6.041	0.000
Postpartum hemorrhage	0.115	0.319	21783	0.062	0.242	916110	-31.613	0.000
Post-birth hospitalization	0.060	0.238	21197	0.051	0.220	867145	-6.195	0.000
Age	33.554	4.206	21783	30.005	5.062	916108	-102.639	0.000
Height	167.347	6.300	21364	166.393	6.350	852598	-21.680	0.000
Weight	68.890	12.163	20375	67.872	13.001	814764	-11.055	0.000
BMI	24.595	4.119	20126	24.509	4.430	800040	-2.727	0.006
Asthma	0.068	0.252	21783	0.068	0.252	916110	-0.044	0.965
Ulcerative colitis	0.011	0.102	21783	0.006	0.076	916110	-9.164	0.000
Epilepsy	0.004	0.063	21783	0.004	0.067	916110	1.024	0.306
Hypertension	0.005	0.070	21783	0.004	0.060	916110	-3.273	0.001
Smoke 1st trimester	0.042	0.201	21408	0.096	0.295	858730	26.692	0.000
Smoke 3rd trimester	0.024	0.153	19334	0.066	0.248	758490	23.439	0.000
Education	4.697	1.319	21741	4.456	1.402	910516	-25.111	0.000
Labor income	65.409	41.438	21755	42.538	33.165	906747	-99.866	0.000
Sickness benefits	1.969	5.376	21755	1.401	4.063	906747	-20.216	0.000
Parental benefits	2.143	4.504	21755	4.515	5.919	906747	58.685	0.000
Maternal health index	-0.249	1.260	21783	0.002	0.997	916110	36.555	0.000
Maternal labor index	0.179	1.096	21765	0.001	0.999	915012	-26.017	0.000
Panel B: Child outcomes and characteristics								
Apgar score	9.630	0.973	21513	9.731	0.810	907738	17.986	0.000
Apgar score ≤ 7	0.018	0.133	21513	0.012	0.108	907738	-8.355	0.000
Birth weight	3,291.359	712.527	21689	3,533.785	586.532	913011	59.831	0.000
Gestational age (weeks)	38.616	2.656	21766	39.313	1.955	915126	51.515	0.000
Head circumference	34.519	1.982	20412	34.928	1.682	872547	34.133	0.000
Length (centimeters)	49.601	3.212	21162	50.378	2.572	903283	43.145	0.000
Gender (male)	0.515	0.500	21765	0.514	0.500	915104	-0.028	0.978
Breech presentation	0.093	0.291	21766	0.038	0.191	915126	-41.754	0.000
Malformation	0.045	0.207	21766	0.035	0.185	915126	-7.311	0.000
Infant mortality rate	4.545	67.264	21783	2.713	52.012	916110	-5.099	0.000
Under 5 mortality rate	5.601	74.630	21783	3.433	58.491	916110	-5.367	0.000
Hospitalization ages 0-1	0.223	0.416	21766	0.164	0.370	915126	-23.131	0.000
Hospitalization ages 1-4	0.147	0.354	21766	0.131	0.338	915126	-6.553	0.000
Child health index	-0.276	1.242	21783	0.004	0.995	916110	40.737	0.000

Note to Table 1. 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 and non-IVF births for the time period 1998-2007. Mean values, standard deviations, observations, t-test with related p-values are displayed.

Table 2: Summary statistics, twins and singletons

	Twins				*Singletons*			
	IVF	*Non-IVF*	*Difference*		*IVF*	*Non-IVF*	*Difference*	
	Mean (1)	Mean (2)	T-test (3)	P-values (4)	Mean (5)	Mean (6)	T-test (7)	P-values (8)
Apgar score	9.429	9.476	1.796	0.072	9.651	9.733	8.157	0.000
Apgar score below 7	0.033	0.033	-0.115	0.909	0.015	0.011	-3.317	0.001
Stillbirth	0.006	0.006	-0.020	0.984	0.003	0.003	-0.356	0.722
Birth weight	2,563.517	2,597.222	2.450	0.014	3,470.749	3,571.744	13.813	0.000
Gestational age (weeks)	36.171	36.239	1.034	0.301	39.237	39.418	7.609	0.000
Head circumference	33.158	33.146	-0.243	0.808	34.814	34.947	6.194	0.000
Length (centimeters)	46.716	46.776	0.773	0.440	50.301	50.506	6.470	0.000
Gender (male)	0.499	0.505	0.522	0.602	0.518	0.515	-0.438	0.661
Breech presentation	0.266	0.268	0.269	0.788	0.050	0.031	-8.350	0.000
Malformation	0.049	0.042	-1.735	0.083	0.043	0.034	-3.543	0.000
Infant mortality rate	13.016	14.852	0.712	0.476	3.873	2.667	-1.821	0.069
Under 5 mortality rate	14.132	16.450	0.857	0.392	5.487	3.484	-2.645	0.008
Hospitalization ages 0-1	0.393	0.367	-2.493	0.013	0.194	0.162	-6.849	0.000
Hospitalization ages 1-4	0.166	0.154	-1.528	0.126	0.149	0.136	-2.996	0.003
Child index	-1.066	-1.021	1.298	0.194	-0.113	0.023	11.000	0.000
N	2,689	10,638			6,197	417,656		

Note to Table 2. 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 and non-IVF births, for the pre-treatment period 1998-2002. Mean values, t-tests and p-values are presented for singleton and twin births.

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)
Mother weight		0.000***		0.000***
		(0.000)		(0.000)
Mother height		0.000***		0.000***
		(0.000)		(0.000)
Smoking 1st trimester		0.003***		-0.001
		(0.001)		(0.001)
Native		-0.000		0.004***
		(0.001)		(0.001)
Labor income		0.000		0.000*
		(0.000)		(0.000)
Sickness benefits		0.000		0.000
		(0.000)		(0.000)
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

Note to Table 3. 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, 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). Estimated date of conception fixed effects are included in all regressions. Standard errors are clustered on the 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 35-39	(8) Ages > 39
IVF × postSET	-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)
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
IVF × postSET	-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)
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

Note to Table 4. 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 twin birth 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 on the 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)
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. 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 index (columns 1 and 4), maternal health index (columns 2 and 5), maternal labor market index (columns 3 and 6). Panel A presents estimates for the full sample and Panel B 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 on the mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 6: 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-stat	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 6. 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 labor market outcomes (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 on the mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 7: Effects of SET on child health

Panel A: *Full sample*

	(1) Apgar score	(2) Apgar score <7	(3) Birth weight	(4) Birth lenght	(5) Head circumfernce	(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)
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
Control mean	9.730	0.012	3530.472	50.367	34.922	39.303	0.514	2.750	3.478	0.039	0.165	0.131	0.036
Control sd	0.813	0.108	589.322	2.584	1.687	1.968	0.500	52.371	58.871	0.193	0.371	0.338	0.185

Panel B: *First-time mothers*

	(1) Apgar score	(2) Apgar score <7	(3) Birth weight	(4) Birth lenght	(5) Head circumfernce	(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)
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
Control mean	9.671	0.015	3434.312	50.143	34.744	39.309	0.515	2.767	3.529	0.051	0.168	0.132	0.038
Control sd	0.898	0.123	588.440	2.687	1.774	2.127	0.500	52.527	59.304	0.220	0.373	0.339	0.192

Note to Table 7. 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. 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). Both FDR and Bonferroni corrected p-values are reported in addition to the conventional. Standard errors are clustered on the mother. * p<0.1, ** p<0.05, *** p<0.01.

Table 8: 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
Control mean	0.007	0.042	0.003	0.009	0.112
Control sd	0.086	0.200	0.052	0.095	0.315

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
Control mean	0.010	0.054	0.003	0.011	0.124
Control sd	0.097	0.226	0.059	0.106	0.330

Note to Table 8. 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). Both FDR and Bonferroni corrected p-values are reported in addition to the conventional. Standard errors are clustered on the mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 9: Effects of SET on maternal health

	Panel A: *Full sample*				
	(1) Emergency C-section	(2) Hemorrhage	(3) Maternal sepsis	(4) Hospital re-admission	(5) Hypersention
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)
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
Control mean	0.084	0.063	0.002	0.051	0.004
Control sd	0.278	0.242	0.039	0.220	0.060

	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)
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
Control mean	0.122	0.072	0.002	0.053	0.003
Control sd	0.327	0.259	0.047	0.225	0.054

Note to Table 9. 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 maternal health outcomes. 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). Both FDR and Bonferroni corrected p-values are reported in addition to the conventional. Standard errors are clustered on the mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 10: Effects of SET on labor market outcomes

Panel A: *Full sample*

	(1) Sickness benefits	(2) Labor income	(3) Parental benefits
postSET*IVF	-0.436***	0.085**	-0.034*
	(0.057)	(0.037)	(0.017)
IVF	1.269***	0.014	0.072***
	(0.043)	(0.030)	(0.013)
FDR p-value (Treat)	0.000	0.297	0.744
Bonferroni p-value (Treat)	0.000	1.000	1.000
R-Squared	0.105	0.276	0.088
Observations	936777	936777	936777
Mean of dep. var.	5.102	10.027	10.155
Control mean	5.094	10.017	10.152
Control sd	4.204	3.242	1.236

Panel B: *First-time mothers*

	(1) Sickness benefits	(2) Labor income	(3) Parental benefits
postSET*IVF	-0.460***	0.051	0.008
	(0.067)	(0.042)	(0.021)
IVF	1.337***	0.137***	0.050***
	(0.050)	(0.032)	(0.016)
FDR p-value (Treat)	0.000	0.297	0.744
Bonferroni p-value (Treat)	0.000	1.000	1.000
R-Squared	0.122	0.262	0.111
Observations	413654	413654	413654
Mean of dep. var.	5.127	10.170	10.191
Control mean	5.115	10.157	10.187
Control sd	4.191	2.977	1.286

Note 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. Each column presents a separate OLS regression with DiD estimates of the impact of the SET reform on maternal labor market outcomes. 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). Both FDR and Bonferroni corrected p-values are reported in addition to the conventional. Standard errors are clustered on the mother. * p<0.1, ** p<0.05, *** p<0.01.

Table 11: 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
IVF*postSET	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 11. 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 on the mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 12: 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 12. 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 on the mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 13: 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)
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)
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

Note to Table 13. 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 on the mother. * p<0.1, ** p<0.05, *** p<0.01.

Table 14: 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 14. 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 on the mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 15: Gelbach decomposition

	(1) Child health index	(2) Maternal health index	(3) Maternal labor index	(4)
postSET*IVF	$\Gamma^{childhealth} = 0.189$***	$\Gamma^{maternalhealth} = 0.032$	0.099***	
			(0.016)	
$\beta_{labor}^{childhealth} = childhealthindex$			0.029***	
			(0.001)	
$\beta_{labor}^{maternalhealth} = maternalhealthindex$			0.050***	
			(0.001)	
$\Gamma_{labor}^{childhealth} \times \beta_{labor}^{childhealth}$				$\delta_{labor}^{childhealth} = 0.005$***
				(0.001)
$\Gamma_{labor}^{maternalhealth} \times \beta_{labor}^{maternalhealth}$				$\delta_{labor}^{maternalhealth} = 0.002$
				(0.001)
Total explained difference				0.007***
				(0.001)

Note to Table 15. 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 on the 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 16: 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

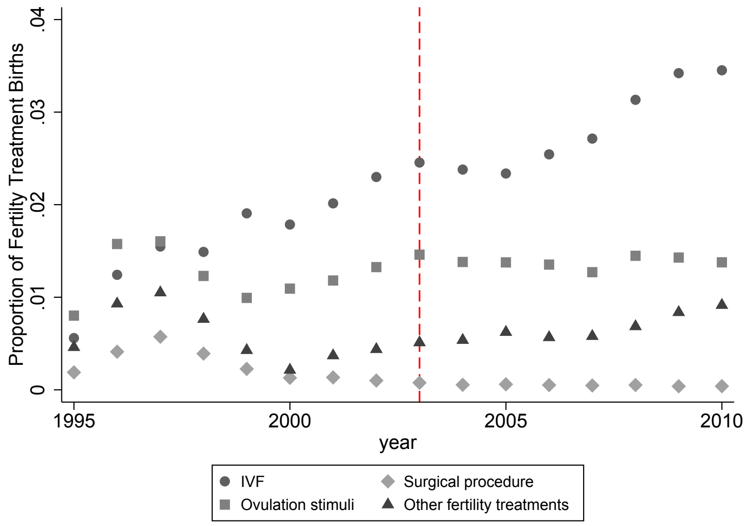
	Panel C: *Removing 2005-2007*			
	(1) Twin birth	(2) Child health index	(3) Maternal health index	(4) Maternal labor index
postSET*IVF	-0.150***	0.154***	0.044*	0.067***
	(0.009)	(0.024)	(0.025)	(0.019)
IVF	0.267***	-0.357***	-0.170***	-0.114***
	(0.006)	(0.016)	(0.015)	(0.013)
R-Squared	0.076	0.021	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 16. 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 on the mother. * p<0.1, ** p<0.05, *** p<0.01.

Appendices

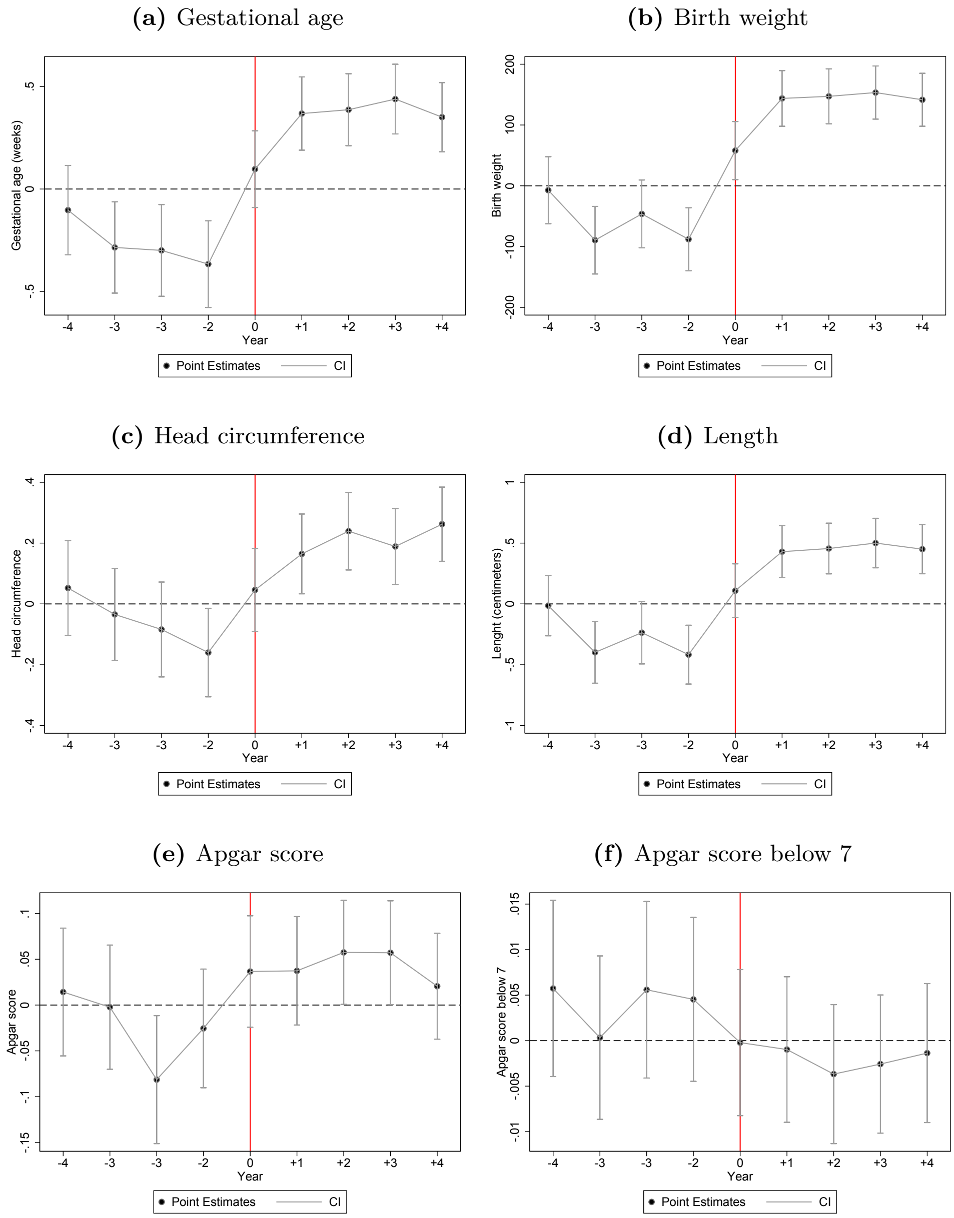
A Figures and tables

Figure A1: ART treatments



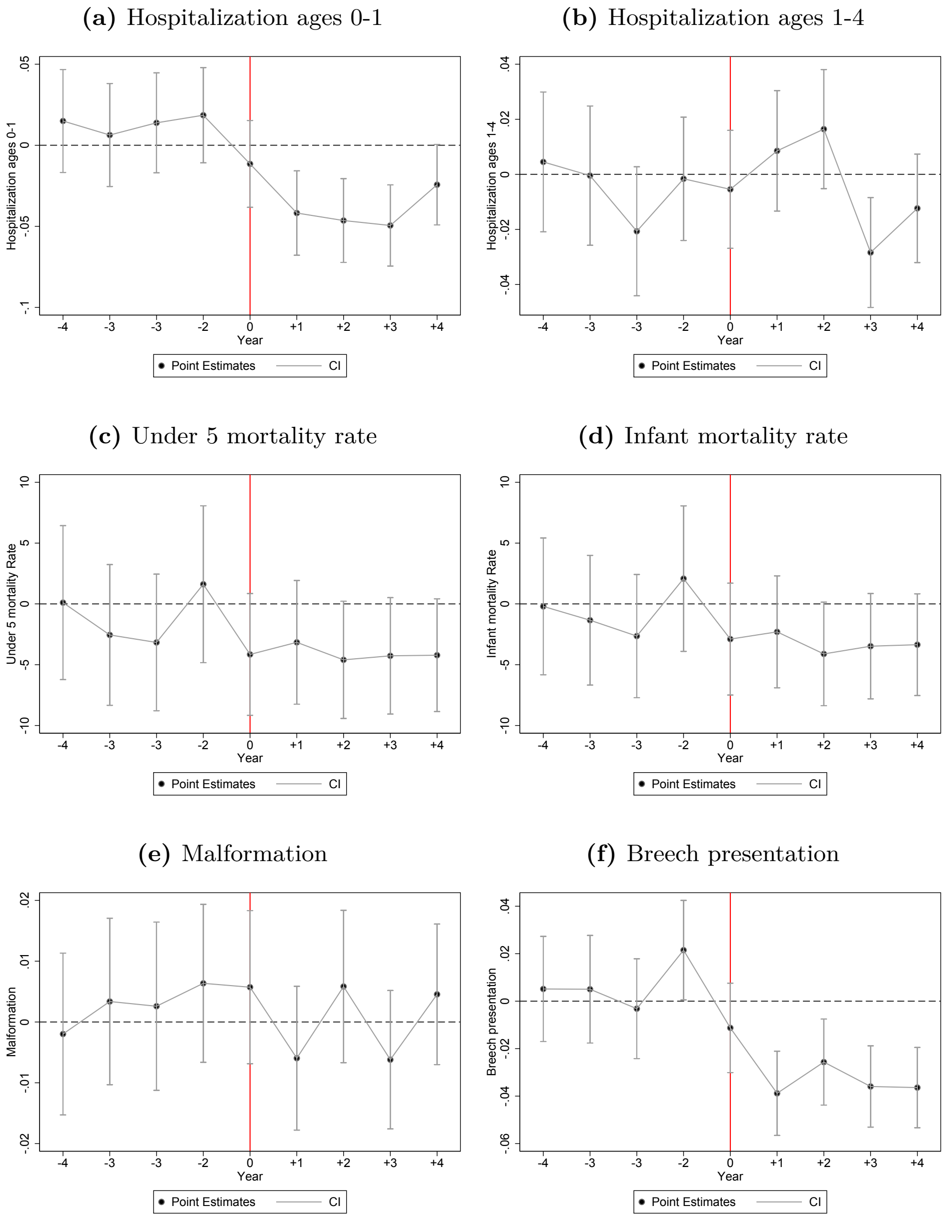
The data are obtained from the Swedish Medical Birth Registry. Trends in different ART treatments are presented in Figure A1. The red-vertical line represents the year of the SET reform.

Figure A2: Child health outcomes



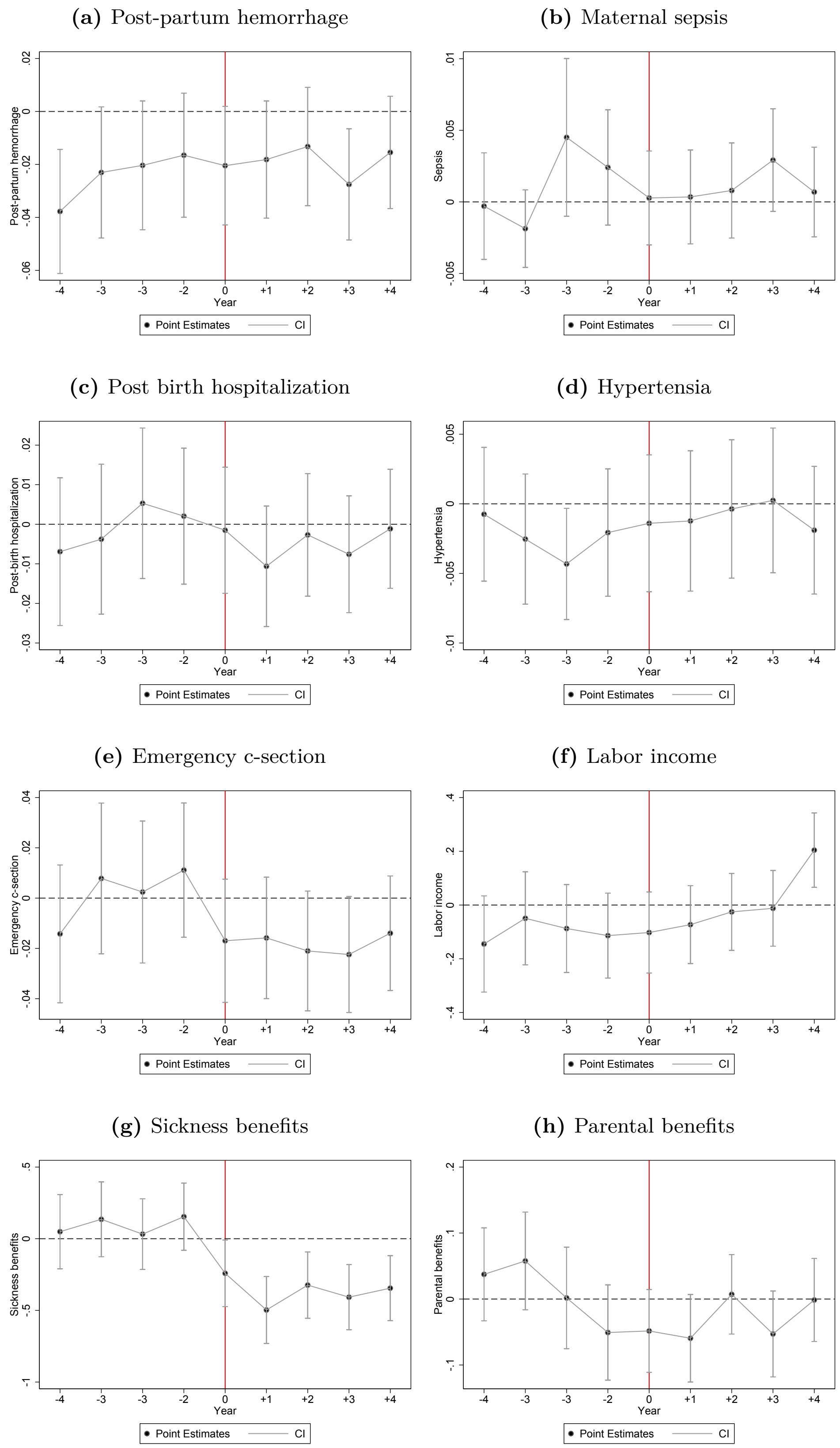
The data are obtained from the Swedish Medical Birth Registry, Swedish National Patient Registry and the Longitudinal integration database for health insurance and labor market studies for the time period 1998-2007. Each figure presents coefficients of interactions between each year and IVF births. The red-vertical line represents the year of the SET reform using the previous year as the omitted category. A full set of maternal controls and fixed effects are included in all regressions (as described in Table 3). Standard errors are clustered on the mother.

Figure A3: Child health outcomes



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 figure presents coefficients of interactions between each year and IVF births. The red-vertical line represents the year of the SET reform using the previous year as the omitted category. A full set of maternal controls and fixed effects are included in all regressions (as described in Table 3). Standard errors are clustered on the mother.

Figure A4: Maternal health and labor outcomes



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 figure presents coefficients of interactions between each year and IVF births. The red-vertical line represents the year of the SET reform using the previous year as the omitted category. A full set of maternal controls and fixed effects are included in all regressions (as described in Table 3). Standard errors are clustered on the mother.

Table A1: 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
Obs	11	11	11
Mean of dep. var.	0.029	0.244	11975.636

Note to Table A1. Aggregated 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.

Table A2: Summary statistics, before-after comparison

	(1) Before SET reform (Jan 1998 - Dec 2002)	(2) After SET reform (Jan 2003 - Dec 2007)	(3) T-test
Twin birth	0.303	0.128	32.426
	(0.459)	(0.334)	[0.000]
Emergency C-section	0.173	0.154	3.641
	(0.378)	(0.361)	[0.000]
Maternal sepsis	0.003	0.003	0.700
	(0.059)	(0.054)	[0.484]
Postpartum hemorrhage	0.111	0.118	-1.550
	(0.314)	(0.322)	[0.121]
Post-birth hospitalization	0.065	0.057	2.349
	(0.247)	(0.232)	[0.019]
Hypertension	0.003	0.006	-3.086
	(0.056)	(0.078)	[0.002]
Labor income	42.432	53.064	-18.401
	(36.319)	(45.336)	[0.000]
Sickness income	4.390	2.919	18.153
	(6.857)	(5.092)	[0.000]
Parental income	17.858	19.448	-13.297
	(8.023)	(9.096)	[0.000]
Apgar score	9.585	9.660	-5.588
	(1.019)	(0.939)	[0.000]
Apgar below 7	0.021	0.016	2.446
	(0.142)	(0.126)	[0.014]
Birth weight	3197.993	3355.470	-16.090
	(754.170)	(675.004)	[0.000]
Gestational age (weeks)	38.309	38.827	-14.225
	(2.890)	(2.459)	[0.000]
Head circumference	34.348	34.631	-9.961
	(2.036)	(1.938)	[0.000]
Length (centimeters)	49.269	49.824	-12.365
	(3.376)	(3.077)	[0.000]
Gender (male)	0.512	0.516	-0.622
	(0.500)	(0.500)	[0.534]
Breech presentation	0.115	0.078	9.270
	(0.319)	(0.269)	[0.000]
Malformation	0.045	0.045	-0.009
	(0.207)	(0.207)	[0.992]
Infant mortality rate	6.640	3.101	3.817
	(81.218)	(55.607)	[0.000]
Under 5 mortality rate	8.103	3.877	4.109
	(89.654)	(62.146)	[0.000]
Hospitalization ages 0-1	0.254	0.201	9.219
	(0.435)	(0.401)	[0.000]
Hospitalization ages 1-4	0.154	0.141	2.637
	(0.361)	(0.348)	[0.008]
Maternal health index	-0.257	-0.244	-0.715
	(1.229)	(1.282)	[0.474]
Maternal labor index	-0.401	-0.190	-12.399
	(1.351)	(1.154)	[0.000]
Child health index	-0.057	0.342	-26.882
	(1.090)	(1.070)	[0.000]
Observations	8,886	12,897	

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. The sample includes IVF births, for the time period 1998-2007. Mean values with standard deviations below, t-tests and p-values are presented for the pre-reform period January 1998-December 2002 (column 1) and post-reform period January 2003-December 2007 (column 2).

Table A3: 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 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. 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 on the mother. * p<0.1, ** p<0.05, *** p<0.01.

Table A4: 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 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. 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 on the mother. * p<0.1, ** p<0.05, *** p<0.01.

Table A5: 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 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. The sample includes IVF mothers for the time period 1998-2007. Mean values, t-test and p-values for the t-tests are displayed.

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:

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

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:

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

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 misclassified 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:

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

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.