Health and Labor Market Impacts of Twin Birth: Evidence from a Swedish IVF Policy

Sonia Bhalotra* Damian Clarke† Hanna Mühlrad‡ Mårten Palme§

October 20, 2022

Abstract

In vitro fertilization (IVF) has allowed women to delay birth and pursue a career, but it massively increases the risk of twin birth. We investigate the extent to which having twins hampers women's careers after birth. To do this, we leverage a single embryo transfer (SET) mandate for IVF procedures implemented in Sweden in 2003, following which the share of twin births showed a precipitous drop of 70%. Linking birth registers to hospitalization and earnings registers, we identify substantial improvements in women's earnings following IVF birth, alongside improvements in maternal and child health and an increase in subsequent fertility. We provide the first comprehensive evaluation of SET. This is relevant given the secular rise in IVF births and the broader rise in the risk of twin birth.

Keywords: twins, IVF, single embryo transfer, career costs of children, gender wage gap, fertility, maternal health, neonatal health, gender
JEL Codes: J13, I11, I12, I38, J24.

---

*Department of Economics, University of Warwick, CEPR, IEA, IZA, CAGE, IFS

†Department of Economics, University of Chile and IZA

‡Institute for Evaluation of Labour Market and Education Policy (IFAU)

§Department of Economics, Stockholm University

*We are grateful to Petra Persson, Maya Rossin-Slater, Neale Mahoney, Shelly Lundberg, Marianne Bitler, Marcus Eliasson, Caroline Hall, Mircea Trandafir, Petter Lundborg, Arizo Karimi, Andreea Mitrut, Randi Hjalmarsson, Sverker Ek, Hannes Schwandt, Eva Meyerson-Milgrom and seminar audiences at the AshEcon conference 2022, 3rd IZA Workshop on Gender and Family Economics, SOLE 2021 Conference, Barcelona GSE Summer Forum for Policy Evaluation in Health 2021, The Institute for Evaluation of Labour Market and Education Policy (IFAU), and the Departments of Economics at Stanford, Bergen, Lund, Uppsala, Gothenburg and Essex for useful comments and discussions. We are grateful to Daniel Pailañir for outstanding research assistance. Sonia acknowledges part-funding for her time from the University of Warwick and support from CAGE. An earlier version of this paper using a shorter data set and studying fewer outcomes appeared in July 2019 as IZA DP No. 12490: Multiple Births, Birth Quality and Maternal Labor Supply: Analysis of IVF Reform in Sweden.

1 Introduction

About 1.6 million twins are born each year worldwide, with one in every 42 children born a twin. The share of twins in all births is rising– records from more than a hundred countries indicate that the global twin birth rate has risen by a third over the past 40 years (Monden et al., 2021). One important driver of this is the steadily increasing use of assisted reproductive technologies such as in vitro fertilization (IVF), some of which reflects delayed parenthood and some of which reflects a rise in infertility.[1] The common practice in IVF is to implant two embryos to increase the chances of success and, as a result, the risk of twin birth with IVF is 10 to 20 times the risk without (Karlström and Bergh, 2007; Kalra and Barnhart, 2011). There are at least two reasons that twin births are also on the rise among unassisted births. One is postponement of parenthood – although the chances of conception fall with age, the probability of having twins increases because older women have higher levels of the follicle stimulating hormone (Beemsterboer et al., 2006; Pison and D'Addato, 2006).[2] The other is trend improvements in maternal health. Conditional upon conception, healthy women are more likely to successfully take a twin pregnancy to term (Bhalotra and Clarke, 2019).[3] Overall, the rising trend in twin birth reflects both women's reproductive health and their pursuit of a career.

In this paper we consider the *consequences* of twin birth for women's reproductive health outcomes and their careers. We leverage the quasi-experimental variation in twin birth rates generated by a reform to IVF procedures implemented in Sweden in January 2003. The policy reform mandated that the default procedure for IVF births should involve a single embryo transfer (henceforth, SET). The SET mandate was driven by scientific research which showed that, on account of advances in IVF technology, pregnancy success rates with single embryo transfers were not that different from success rates with double embryo transfers (DET) (Lukassen et al., 2005; Karlström and Bergh, 2007; Lundin and Bergh, 2007; Criniti et al., 2005; Kutlu et al., 2011). The abrupt introduction of SET disrupted reproductive outcomes for women using IVF. The twin birth rate

[1] A meta-analysis of studies of sperm counts of men (unselected on fertility) in western countries concludes that it has halved in the last 40 years (Levine et al., 2017). Couple-level infertility, defined as failure to achieve pregnancy after 12 months or more of regular unprotected sexual intercourse, is currently estimated at between 7% and 15% of all couples (Geyter, 2021; Ekechi, 2021). The growth in fertility treatments possibly also reflects their increasing success in line with improvements in IVF technology; falling costs as governments expand subsidies for IVF; and growing information and acceptability. A third of US adults say they have used fertility treatments or know someone who has (Pew Research Center, 2021).

[2] In a 2015 publication, the German Federal Institute for Population Research reported that the percentage of women giving birth at age 35 or older had risen from 7.6% in 1981 to 25.9% in 2015. OECD figures indicate that the average age at first birth has risen by about 7 years since the 1970s, and 3.4 years since 2000 (OECD, 2021).

[3] Bhalotra and Clarke (2019) analyse close to 17 million births in 72 countries, of which 2.73% are twins. They show that mothers of twins are selectively healthy. This is the case in richer and poorer countries, and it holds for sixteen different markers of maternal health, including health stocks and health conditions prior to pregnancy (height, obesity, diabetes, hypertension, asthma, kidney disease, smoking), exposure to unexpected stress in pregnancy, and measures of the availability of medical professionals and prenatal care. The effects are sizable, with a 1 standard deviation improvement in the indicator increasing the likelihood of twinning by 6% to 12%. The authors provide evidence of selective miscarriage being the key mechanism, using U.S. Vital Statistics data for 14 to 16 million births.

fell sharply from about 30% pre-SET to 5% soon after. By 2018, it had fallen to 2.54%, which is one of the world's lowest twin birth rates among IVF users.

The twin birth rate among IVF users continues to be 25%–30% in many countries (National Board of Health and Welfare, 2020). Moreover, the share of women using IVF shows a steady rise. The share of all births owing to IVF now exceeds 3% in many industrialized countries (de Mouzon et al., 2010) and it is estimated to be growing at about 7% per annum. These facts underline the relevance of our analysis. Following the lead of Sweden, other countries including Belgium and Turkey have mandated SET. The advantages of SET are increasingly recognized, and many countries actively encourage elective uptake of SET.[4] Nevertheless, multiple (typically double) embryo transfers are still prevalent in many countries.

Our results contribute to policy decisions over SET, providing the first evidence of how it impacts women's earnings and fertility. Since the SET mandate acts to lower the risk of twin birth, our results are also relevant to understanding impacts of twin birth in the wider population of women with unassisted births. Ours is the first study to account for selection into twin birth (which Bhalotra and Clarke (2019) show is pervasive) by exploiting the SET rule as a natural experiment. A large literature has imposed the strong assumption that twin births are exogenous and used this to learn about behaviour, see (Bhalotra and Clarke, 2020, 2022). We delineate our contributions to the literature more comprehensively in the next sub-section. The rest of this section describes the data, the identification and our main findings.

We merge the Swedish medical birth register with administrative data on hospitalization, death and earnings, and link children to mothers and fathers using the multi-generation register. We analyse birth cohorts 1998-2007, a window around the SET reform in 2003, and track outcomes for up to nine years after birth. The estimation sample, including all women (those who do and do not use IVF) contains individual longitudinal data for approximately 895,000 births and 908,000 children. We study multiple indicators of maternal and child health, earnings of mothers and fathers, and fertility continuation. We consistently report p-values adjusted for multiple hypothesis testing.

While implementation of the SET mandate created a sharp drop in twin birth among IVF-users, the share of twin births among unassisted births was stable. This makes it unlikely that the post-SET drop in twin birth among IVF-assisted births is the result of aggregate factors such as improvements in maternal health which impact both IVF and non-IVF births. We control for all such aggregate trends by adopting a double difference approach, comparing the evolution of outcomes before vs after the SET mandate, between IVF and non-IVF births.

The identifying assumption is that outcomes of IVF vs non-IVF users would have evolved similarly in the absence of the SET reform.[5] We investigate this by estimating event study models,

[4]The UK, US and Switzerland are among countries actively encouraging elective SET, see `https://www.hfea.gov.uk/about-us/our-campaign-to-reduce-multiple-births/`; `https://www.cdc.gov/art/patientresources/transfer.html`; and Geyter (2021).

[5]The difference in difference approach does not require that the treated (IVF) and untreated (non-IVF, or

controlling flexibly for the age and education of the mother and father. We find no evidence of differential pre-trends. We nevertheless provide bounds on the estimates obtained by relaxing the parallel trends assumption, following Rambachan and Roth (2020).

If the identifying assumption holds, we identify the causal impact of mandating SET, a policy relevant parameter. However, if we wish to interpret the estimates as identifying impacts of replacing twin with singleton birth then it is relevant to investigate endogenous (SET-driven) selection into IVF treatment. We do this and find it is limited. We nevertheless employ partial identification methods to account for selection on unobservables and, for an appropriately matched sample, we provide estimates conditional on mother fixed effects.[6]

Our main finding is that mandating SET for IVF-users led to a decrease in the career costs of children, alongside significant improvements in maternal and newborn health, and despite an increase in the probability of a subsequent birth (among the 74% of IVF mothers that are at first parity). The identified impacts are fairly pervasive, being evident across age, education and body mass index (BMI) of the mother and, for continuous outcomes (birth weight, gestational age, earnings), across the distribution of the outcome. Overall, SET acted to resolve the trade-off that women face between delaying birth to establish careers with the fallback of seeking IVF treatment, and the elevated risks from IVF treatment of suffering perinatal complications, unwanted (multiple) births, and post-birth earnings penalties. Our estimates show that the benefits accruing from reduced hospital costs, increased earnings of women and and increased potential earnings of children vastly overwhelm the costs associated with the SET reform.

We now elaborate our findings by outcome domain. This showcases the richness of the data, and it helps position the contributions of the paper against the available evidence (also see the next section). Neonatal outcomes for children born of IVF show a dramatic improvement across a range of indicators after the SET mandate, the baseline gap relative to non-IVF births narrowing by 50 to 75 percent.[7] We identify increases in gestational age, birth weight and APGAR score,

---

mothers with natural births) exhibit balance in levels, only in trends. In other words, identification does not require that there is no selection into IVF, but rather that this selection does not systematically change when the SET reform is put in place. Women undertaking IVF tend to be older and exhibit healthier pregnancy behaviours, and both of these factors predict twinning irrespective of SET vs DET procedures (Bhalotra and Clarke, 2019). By leveraging the SET mandate we are able to isolate policy-driven variation in twinning from the variation predicted by characteristics.

[6]These formal tests aside, looking at the likely direction of any selection bias, and at the data mitigates selection concerns. First, if women with more favourable characteristics were more likely to undertake IVF following the SET mandate then this would lead to higher twinning alongside better health and earnings (Bhalotra and Clarke, 2019), while the SET reform led to lower twinning alongside better health and earnings. Second, the proportion of women seeking IVF does not show a discontinuous change at the date of the SET reform. If it did, this might signal that women at the margin are different and hence that there is a compositional shift. Third, there is evidence from Denmark that IVF success rates are idiosyncratic, showing no association with a rich set of mother characteristics (Lundborg et al., 2017). So even if SET changes success rates per attempt (albeit not cumulatively), this should not trigger a compositional change. Finally, we provide evidence from the National Board of Health in Sweden which indicates that the SET mandate did not change cumulative IVF success rates.

[7]We do not expect full convergence. Only 70% of IVF births were conceived with SET after the reform

and declines in hospitalization and an index of 17 neonatal morbidities. The probability of low birth weight at the commonly used medical thresholds of 1500 and 2500 grams fall by 1.2 and 7.8 percentage points (pp) respectively. Maternal morbidities, maternal hospitalization and C-section rates also fall after SET. The effects are large, for example, the number of nights mothers spend in hospital after birth falls by 0.63, closing 63% of the baseline excess over non-IVF births, and receipt of sickness benefits declines by 8.2%, eliminating 40% of the excess relative to non-IVF births.

We find that women using IVF are 7.2 pp more likely to have another birth after SET, driven by women at first parity, for whom the increase is 10 pp. In Sweden, on average about half of all women have a second child. Before the SET reform women using IVF were less likely than other women to proceed to a subsequent birth– 46% compared with 53% of women and this is likely, at least in part, because a larger share of IVF users had twins. The SET reform, by virtue of reducing twin birth risk among IVF-users, closes more than two-thirds of this gap. The increase translates to 0.11 additional births among IVF women which occur within 3 years of the index birth.

Averaged over the nine years after birth, the earnings of women giving birth with IVF assistance are 5.6% higher after SET than before SET. We find no discernible impact of the SET mandate on father's earnings. It follows that SET produced an increase in the long run earnings of mothers relative to fathers, a measure of the child penalty that Kleven et al. (2019) argue is an increasingly important explanation of the gender earnings gap. This differential has also been used as a marker of women's bargaining power in the household that impacts their consumption, autonomy, and risk of domestic violence victimization (McElroy and Horney, 1981; Lundberg et al., 1997; Aizer, 2010).

The post-SET increase in mother's earnings is evident at the extensive and intensive margins of both fertility and earnings. Earnings increase for women having their first birth (extensive margin), as well as for women at higher parity (intensive margin). The increase is larger for women at second or higher parity, consistent with these women being less likely to want an additional birth. We find significant increases in employment (extensive margin) and earnings (intensive margin) but the latter dominates as women's labour force participation rate in Sweden is high and it is uncommon that women drop out after birth.

Our results illuminate the trade-offs that arise in postponing birth and undertaking IVF, while also demonstrating the potential for recent advances in IVF technology to mitigate these trade-offs. These trade-offs are likely smaller in Sweden than in most other countries because Sweden provides universal health care to a high standard, and unusually generous parental leave. As a result, SET mandates for IVF are likely to have larger effects in other countries than we identify in Sweden.

---

because the mandate allowed exceptions, for example, for older women (see section 2).

1.1 Existing literature and contributions

This sub-section contextualizes this paper with respect to related research delineating its contributions to research on IVF and more generally research on fertility and twin birth.

Our first contribution is that we comprehensively document impacts of the SET-led shift away from twin birth among IVF-users on a rich set of indicators of neonatal and maternal health. Biomedical studies have already documented that SET, compared to DET, in the Swedish context, led to lower twin birth rates and improved maternal and neonatal health (Karlström and Bergh, 2007; Lundin and Bergh, 2007; Saldeen and Sundström, 2005; Sazonova et al., 2011; Thurin et al., 2004), which lowered the costs associated with IVF (Lukassen et al., 2005; Kjellberg et al., 2006). Our contribution relative to the biomedical literature is primarily to consider additional outcomes, in particular, fertility and earnings, and to look at several outcomes over a longer period (nine years from birth).

In the economics literature, Bitler (2008) uses variation in access to infertility treatments induced by state-level insurance mandates in the US to find that *expanding access to IVF* leads to an increase in twin birth and hence an increase in the healthcare costs of birth. In contrast, we estimate impacts of regulating the number of embryos transferred *conditional upon IVF use*, which results in a decrease in twin births and a decrease in healthcare costs of birth. The distinction is important because our estimates not only document impacts of twin birth but show how a switch to SET can largely mitigate these impacts.[8]

Our second contribution is that we model impacts of the SET reform on fertility continuation. This provides (ITT) estimates of how the occurrence of twins vs singletons at first birth modifies the demand for a subsequent child. Our finding that mandating SET (and thus reducing the chances of a twin birth) led to higher subsequent fertility among IVF users is consistent with some women wanting two children. For these women, SET results not in a change in fertility but instead in birth spacing (while for women who do not proceed to have another birth, there is also a change in fertility). There is relatively little causal evidence of the impacts of birth spacing vs additional births on women's labour market outcomes, though see Karimi (2014); Adda et al. (2017); Bhalotra et al. (2022).

We provide estimates of the excess motherhood penalty exerted by twins in comparison with singleton births, using a policy experiment that is uniquely positioned to identify this object. In doing this, we contribute to the literature on the career costs of children (Adda et al., 2017; Kleven et al., 2019; Berniell et al., 2021). It is relevant to distinguish our contribution from the twin instrument literature. In this literature, fertility is instrumented with an indicator for twin birth, and the estimates are interpreted as reflecting the impact of having an additional child, largely ignoring direct impacts of twin birth arising from their closer (close to zero) birth spacing and

[8] An additional distinction of our work is that we investigate not only child health but also maternal health, and we use a richer set of indicators of child health, and that track child hospitalizations through to nine years after birth.

their weaker birth endowments (Cáceres-Delpiano, 2006; Conley and Glauber, 2005; Rosenzweig and Zhang, 2009; Qian, 2017). In this paper, we provide ITT estimates of having a twin birth as compared with having a singleton birth.

Our estimates are averaged over nine years after birth, a period during which we allow that additional births occur and that these additional births also influence a woman's earnings trajectory. Allowing for endogenous fertility continuation is important because some women giving birth after the SET mandate possibly wanted two births. Our estimates further contribute to identifying impacts of extensive margin fertility on women's earnings, i.e. impacts of a first birth, see Lundborg et al. (2017).[9]

Overall, we substantially extend the evidence of impacts of twin birth on maternal and child health, and provide new evidence of impacts of twin birth on future fertility and earnings.[10] Although it is widely acknowledged that availability of IVF and other assisted reproductive technologies has significantly enhanced women's choice set, slackening the career-family tradeoff by allowing women to invest in their careers and delay birth (Gershoni and Low, 2021a,b), there is surprisingly limited evidence of whether the substantial increase in the risk of twinning associated with IVF compromises the careers of women *after* birth.

Our second contribution is to provide a comprehensive evaluation of mandating SET for IVF treatments, along with a crude benefit-cost assessment. Our assessment includes short run benefits (dominated by health gains) and longer run benefits (dominated by earnings gains for both the mothers and the children), and we account for the marginal cost of public funds. While this is only a back of the envelope calculation and the division of costs between the public and the private purse is different across countries, it does serve to make salient the range of dynamic benefits flowing from the SET mandate.[11]

The rest of this paper is structured as follows. Section 2 describes IVF and the SET reform in Sweden. Sections 3 and 4 discuss the data and empirical strategy. Section 5 presents the reform impacts on fertility, child and maternal health and 6 present results related to parental income and the career costs of children. Section 7 then documents robustness to alternative methods, and Section 8 concludes.

[9] Lundborg et al. (2017) use IVF-success within a sample of IVF-users in Denmark as an instrument for fertility. In fact, IVF-success switches the number of children from 0 to 1 for about 75% of Danish IVF-users, and from 0 to 2 for the other 25%. Their estimates of the child penalty are thus a weighted average over women who have singleton and twin births. The increment in fertility is, on average, 1.25. Thus, they capture the extensive margin for a birth *event* but not for a birth. We similarly estimate impacts of a birth event, leveraging SET to consider the switch into 1 birth relative to the switch into 2 births.

[10] In fact we provide reduced form estimates of impacts of mandating SET. These need to be scaled by impacts of SET on twinning to deliver estimates for twinning.

[11] In countries like the US with relatively restrictive public coverage for IVF (Bitler and Schmidt, 2012), it is likely to be harder to mandate SET (Karlström and Bergh, 2007; Pinckney-Clark et al., 2016) because when families privately bear a large share of the costs of IVF, they may elect DET even when they do not have a preference for twins because DET is less expensive than two cycles with SET (Hamilton et al., 2018).

2 Background: IVF eligibility and technological advances

The global uptake of IVF has increased steadily over time, no doubt encouraged by improvements in availability of contraception and by both demand and supply forces encouraging women's labour force participation. In 2018, about 8 million children had been born as a result of IVF (European Society of Human Reproduction and Embryology, 2018). In Sweden, in our sample years of 1998-2007, 2.15% of all births occur with IVF assistance.

2.1 IVF treatments in Sweden – access and eligibility

All permanent residents in Sweden have access to heavily subsidized health care offered by both private and public providers. For most medical services, there is a small fee, capped at 1100 SEK (approximately 125 USD) per annum and there are typically no additional costs. Health care is mainly funded by tax revenues, and only 2% of residents have private health insurance (Anell, 2008). Eligibility criteria for subsidized coverage of IVF treatments are as follows. The couple should be in a stable union, either legally married or co-habitating for at least two years, although since 2016 single women are also allowed to access publicly funded IVF treatment and, lesbian couples have been allowed access since 2005. The woman should have no previous children, either biological or adopted. IVF is available for second and higher order births but this is not publicly funded.[12] Finally, a medical assessment of the woman is conducted to confirm that her body mass index (BMI) is within the normal range, that there is no evidence of risky behavior such as smoking and use of alcohol and other drugs/narcotics. Other mental and physical illness and disability are also considered before offering treatment. The suggested maternal age for starting the first treatment is below 40 and the guidelines suggest that any remaining embryos/egg cells should be transferred before age 45. The age of the man should lie between 25 and 56 years. BMI thresholds and age restrictions are county-specific, for example, the maximum age of the mother in Örebro county is 43 while in Norrbotten county it is 37 (Alm, 2010). A couple is allowed three rounds of treatment (follicle aspiration), and any remaining embryos and eggs of good quality are frozen.

2.2 Technological advancements and new guidelines on SET

IVF technology has evolved substantially in the last four decades on account of advances in ovarian stimulation, embryo culture, cryopreservation techniques and genetic testing (Eskew and Jungheim, 2017). This has contributed to improved pregnancy success rates. In Sweden in the 1980s, IVF pregnancy success rates hovered around 5-10%. Multiple embryos were transferred, as many as four embryos at a time, so as to raise the odds of a successful pregnancy (Wikland, 2005). In 1993, all IVF clinics in Sweden implemented a voluntary reduction in the number of embryos rou-

[12] The cost of a private IVF treatment in 2021 is around 80,000 SEK (about 9,153 US$).

tinely transferred, from three to two. This resulted in the virtual elimination of the conception of triplets by IVF, but pregnancy rates and live-birth rates remained essentially unaffected at 35 and 25 percent per transfer, respectively Bergh et al. (2005).

In the late 1990s and early 2000s, biomedical scientists moved to analyse the case for replacing double (DET) with single embryo transfer (SET). They were motivated by the evidence that both the mother and the child suffered significantly higher risks with twin conception (Bergh et al., 2005; McLernon et al., 2010). Observational studies emerged, indicating that a change in protocol from DET to SET would preserve pregnancy success rates and delivery rates (defined as at least one live birth) (Vilska et al., 1999; Gerris et al., 2001; De Sutter et al., 2003; Tiitinen et al., 2003; van Montfoort et al., 2005). Randomized controlled trials (RCTs) suggested a lower pregnancy success rate when comparing one elective SET procedure with one DET procedure (Martikainen et al., 2001; Gerris et al., 1999; Thurin et al., 2004; Gardner et al., 2004; Lukassen et al., 2005), but similar pregnancy rates across procedures (Thurin et al., 2004; Lukassen et al., 2005). The RCT studies addressed the issue of selection that mars observational studies, but were not necessarily more reliable because their external validity is questionable – they typically recruited couples with at least two embryos of good quality Bergh et al. (2005)). Their results may therefore not apply at the population level. A relatively recent meta-analysis however also argues that the cumulative success rates from SET and DET are similar (McLernon et al., 2010).[13]

Provoked by the evidence, the Swedish National Board of Health and Welfare issued new guidelines, on January 1 2003, recommending SET as the default IVF procedure replacing DET. The new guidelines allowed for exceptions for women with a low perceived risk of twinning. In particular, women with low embryo quality, women aged above 38 years and women with more than three previously failed IVF cycles were still allowed DET, provided that they were informed about the potential risks for the mother and child (Saldeen and Sundström, 2005). At the time that SET was implemented, there were no other changes in the IVF treatment procedure with respect to medication, technique or equipment (Saldeen and Sundström, 2005).[14]

[13]SET often involves transferring one fresh embryo and a *subsequent* transfer of a frozen embryo if needed. The success rates of SET converge towards those of DET once one accounts for cumulative success across both procedures. The additional cycle in the SET mode does not require a whole new IVF cycle with egg aspiration and hormonal treatment, and it is hence substantially less costly to perform than the initial SET procedure, both in terms of lower medical costs and less strain on the woman.

[14]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). 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).

2.3 Pregnancy rates and live birth rates in Sweden around the time of the SET mandate

Data from the National Board of Health and Welfare suggest that both pregnancy success rates and live birth rates (of at least one live birth) were no lower following the SET mandate (see SoS, 2013; Bergh et al. (2005)). Using their aggregate data, we illustrate this in Figures A1a and A1b. The solid line in Figure A1a suggests that, before the SET mandate, approximately 20% of all started IVF cycles (which is defined as the process of egg retrieval) resulted in a live birth. There was a slight dip in this rate immediately after the reform, which was restored soon after and, importantly, the yearly post-SET variation is similar to the pre-SET variation. Similarly, the success rate of embryo transfers (which is defined as the process of egg implantation, after fertilization) remains broadly stable at about 25% on both sides of the date of the SET reform.[15] The numbers of IVF cycles started and embryo transfers also remain on their pre-reform trend, see Figure A1b.[16]

3 Data and descriptive statistics

3.1 Data

Data sources We use the the Swedish Medical Birth Registry, provided by the National Board of Health and Welfare, which covers approximately 99% of all births in Sweden. This contains information on pregnancy, delivery and post-partum outcomes. It includes maternal characteristics including age, parity, body mass index (BMI), chronic diseases, tobacco consumption and prenatal conditions and treatments. Information on the father is not in the birth register but we obtain the father's age, education and whether he was born in Sweden from LISA, another register, discussed below. We can identify fathers even if they no longer live with the mother or child. Our indicator for IVF includes standard IVF and IVF with ICSI. We do not observe unsuccessful fertility treatments. We observe IVF births in the register if the treatment results in a successful pregnancy delivered after week 22, and the register flags assisted births including standard IVF, intracytoplasmic sperm injection (ICSI), surgical procedures and ovarian stimulation. In a paper that relies upon this, Lundborg et al. (2017) demonstrate that IVF success in Denmark is uncorrelated with a rich set of observable mother characteristics.

We supplement the birth register with data from other registers, retrieving the link between each mother, father, and child from the multi-generation Register, provided by Statistics Sweden. For children, we obtain additional medical information on hospitalizations during childhood from the National Patient Registry (NPR) and information on child mortality from the Cause of Death

[15]The success rate of about 25% for embryo transfers is higher than the 20% for egg retrieval because a proportion of egg retrieval attempts do not generate viable fertilized eggs which can be implanted, in which case there is no embryo transfer.

[16]Note that, after SET, women may undergo more than one procedure within a given cycle or embryo transfer to achieve the same success rate.

Registry, both provided by the National Board of Health and Welfare. For parents, we obtain administrative data on income and educational attainment from the Longitudinal Integration Database for Health Insurance and Labor Market Studies (LISA) provided by Statistics Sweden. For fathers, we further obtain information on age and whether he was born in Sweden from the total population registry.

Sample The sample consist of all births in Sweden conceived during 1998-2007. The data on post-birth outcomes including hospitalization, mortality and income are available until 2016. As the last cohort in the birth sample (babies born in 2007) can be followed for at most 9 years after birth, we measure post-birth outcomes over a window of 9 years so that we have a balanced panel. So, for example, for a birth that occurs in the year 2000, we track hospitalization and parent earnings up until 2009. We remove triplet and higher order births (532 births). We also remove a small number of observations with missing or conflicting information on birth date or gestational age (1071 births), parental age (367 births), region of birth (374 births), and birth order (16 births). The data include 22,183 IVF births and 932,822 non-IVF births conceived during this period . While the unit of observation for child health is births, the unit of observation for maternal outcomes is pregnancy. Because women older than 38 were exempt from the SET mandate, we primarily focus on women younger than 39 for the main analysis (19,563 IVF births and 888,675 non-IVF births: 16,097 unique IVF mothers and 588,308 unique non-IVF mothers). The share of IVF births in this sample is 2.15%.

Outcome variables *Fertility:* The outcome variables are an indicator for twin as opposed to singleton birth, the number of subsequent birth over the 9 years following the index birth and the extensive margin indicator of this, defined as the probability of having no further births. A birth is an event, so an additional birth event is coded as one subsequent birth, irrespective of whether it results in a singleton or twins.

Maternal health: We also examine maternal morbidity, length of hospital stay during delivery (number of nights hospitalized), and whether the birth was delivered by emergency Caesarean section. Following Wennerholm et al. (2019), we construct a composite index for severe maternal morbidity, which is set to 1 if there was at least one case of morbidity, and zero otherwise, See Table A1.[17]

Neonatal health: We investigate birth weight, gestational age, nights hospitalized during the first nine years of life, the probability that the APGAR score is below 7, severe neonatal morbidity, infant mortality, and under-5 mortality.[18] For some variables we create more than one indicator

[17]The variables included in this composite index are postpartum hemorrhage >1000 ml, hysterectomy, other major surgical intervention (uterine compression sutures such as B-Lynch, uterine artery ligation or embolization, internal iliac artery ligation, or intrauterine balloon tamponade), venous thromboembolism, maternal sepsis, maternal death, 3rd or 4th degree perineal laceration, anesthesia complications, other obstetric injuries, post-partum depression, placenta complications, complications due to multiple birth, uterine rupture, eclampsia, preeclampsia, gestational diabetes mellitus, cervical lacerations, chorioamnionitis, wound infection, endometritis, and urinary tract infection.

[18]APGAR, measured 5 minutes after birth, stands for "appearance, pulse, grimace, activity, respiration"

outcome, for example, we study low birth weight and preterm delivery in line with medically relevant cutoffs. For birth weight and gestational age, we also plot effects across the distribution. We construct a composite index for severe neonatal morbidity using registered medical diagnoses and surgical procedures defined according to the International Classification of Diseases – 10th Revision (ICD-10) and the classification of care measure (Swedish KVÅ-codes), using the criteria laid out in Wennerholm et al. (2019). See Table A2. The index is a binary variable set to 1 if there was at least one case of severe morbidity, and zero otherwise.[19] Summary measures of morbidities by the IVF status of the birth are presented in Tables A3 for maternal morbidity, and A4 for child morbidity.

Parental income: We study wage earnings and income from sick pay benefits. Parental wage earnings represent taxed annual earnings from gainful employment. The public sick pay insurance program replaces 80 percent of forgone earnings below a social security ceiling after two weeks in a work absence spell due to temporary health deficiencies. The first two weeks in a spell are financed by the employer. A certificate from a physician confirming the health deficiency is needed after 10 days in a spell. Both income variables are expressed in annual amounts in Swedish kronor (SEK) using the 2018 consumer price index. We study incomes averaged over a period of 9 years after birth, and provide estimates of the impact of SET across the earnings distribution. Only 8% of all women have zero earnings within 3 years of giving birth, a marker of the high rate of labour force attachment of Swedish women. We present estimates for earnings including the zeroes but we also provide separate estimates excluding them and modelling the probability of shifting from zero to nonzero earnings after SET. Additionally, we examine the impact of SET on women's earnings relative to the earnings of their partners.

Covariates We control for age fixed effects (FE), FE for educational attainment, and whether born outside Sweden for both mother and father, the mother's pregnancy order FE, child gender, region of birth and conception year-month FE. In order to control for maternal health predictors of twinning (Bhalotra and Clarke, 2019) we also control for whether the mother smoked in early pregnancy and her pre-pregnancy BMI.[20] We document that the results are, in general, not sensitive to these controls, and additionally are robust to allowing for trends in maternal and paternal characteristics.

---

and is a five-criterion evaluation method, indicating the general health condition of the newborn baby.

[19]The indicators of severe morbidity are: APGAR score below 4, pneumonia, sepsis, birth trauma (fractures, neurologic injury, retinal haemorrhage or facial nerve palsy), hypoglycemia, plexus injury, stillbirth, birth weight <1500 g, preterm birth before 32 weeks, umbilical artery pH <7.0, hypoxic ischemic encephalopathy, intracranial hemorrhage, neonatal convulsions, meconium aspiration syndrome, mechanical ventilation, cardio-respiratory resuscitation and therapeutic hypothermia.

[20]We control for the highest level of education, 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.

3.2 Descriptive statistics

Summary statistics are provided in Appendix Tables A5 and A6 for parents and children respectively, stratifying by IVF status. Table A7 additionally stratifies by whether the birth is a twin or a singleton.

Refer to Table A5, which summarizes outcomes (Panel A) and characteristics (Panel B) for IVF vs non-IVF users, averaging the data for the baseline (pre-SET) period of 1998-2002. Recall that we only have information on IVF parents who succeed in having a birth. The baseline rate of twin birth in the IVF sample is 19%, compared with less than 1% in the non-IVF sample. In the IVF sample, 74% of all women are first-time mothers, compared with 44% in the non-IVF sample. Women using IVF are older, taller, heavier and less likely to have smoked during and before pregnancy.[21] They have higher education and earnings, consistent with more educated women being more likely to delay birth so as to fulfil career ambitions, also see Table A8, which shows rates of IVF births by age. Despite this but in line with being older and having higher rates of twin births IVF-using women exhibit significantly higher maternal morbidity and birth-related complications, see Tables A3 and A4.

Table A6 reveals the significant extent to which, pre-SET, multiple indicators of in utero and neonatal health are also worse for children born of IVF. Is this entirely accounted for by IVF-children being more likely to be twin births? To assess this we summarize baseline indicators of child health separately for twins and singletons, providing tests of the difference in means between IVF and non-IVF births, see Table A7. For 6 of 8 indicators, IVF-twins are not significantly different from non-IVF twins.[22] However, for 6 of 8 indicators, IVF singletons have significantly worse health than non-IVF singletons. This may reflect, inter alia, that singletons born after an IVF procedure were conceived as twins, with only one surviving to birth. It is relevant here because it tells us that we should not expect complete convergence of IVF outcomes towards non-IVF outcomes after the SET reform.[23]

The fact that women using IVF are older and have healthier pregnancy behaviours predisposes them toward twin birth (Bhalotra and Clarke, 2019). In other words, even under SET, IVF users will have a higher twin birth rate than non-IVF users. We adjust for this by controlling for individual characteristics including age and pregnancy health indicators. Importantly, the SET reform isolates variation in twinning that is driven by the reform, as long as the reform does not generate selection into IVF, a concern that we investigate and allay.

[21]The age distributions in the two samples is plotted in Appendix Figure B1.

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

[23]The singleton sample is much larger than the twin sample and this will contribute to more precise estimation of differences in the singleton sample. However we note that the medical literature (Pinborg et al., 2013; Sazonova et al., 2011) has also noted that IVF singletons have worse birth indicators.

3.3 Trends in IVF and twinning

The share of all births assisted by IVF evolves smoothly after the SET mandate. The proportion of all live births assisted by IVF is in Figure 1a, and the trend is smooth around the date of the SET reform. The share of IVF births will be a function not only of the rate at which women select into IVF, but also of the IVF-success rate. As discussed in the preceding section, the pregnancy success rate among IVF users was maintained at about one-quarter following the SET mandate (Karlström and Bergh, 2007). Also see Figure A1a, which suggests no significant change in deliveries per cycle and per embryo transfer, and Figure A1b, which shows that the number of IVF treatments performed is smooth around the cut-off. Trends in the share of births achieved with each type of ART procedure are presented in Figure B2, which shows that IVF is the only ART procedure exhibiting a trend. This limits concern about endogenous shifts in the composition of mothers using IVF after SET, but we nevertheless investigate this in Section 7.1.

The share of IVF births conceived with SET rises sharply after the SET mandate. This is a measure of compliance with the SET reform. The trend in the proportion of IVF births achieved using SET rather than DET is in Figure 1b, showing a sharp increase coincident with the SET reform.[24] This share is stable at about 10% for the decade to the year 2000, increasing to 30% in 2002, following which there is a significant jump to 60% in 2003. It continues to rise to 70% in 2005, after which it stabilizes. The share does not rise to 100% because of the exemptions we outlined in Section 2, that allowed some IVF procedures to proceed with DET. In Section 7.1, we discuss the elective increase in the share of births conceived with SET during 2001-2002, in advance of the mandate, and that one region, Skåne, introduced the reform in 2001.

The share of twins among IVF births shows a sharp drop after the SET mandate. Figure 1c shows a sharp drop in the share of twin births among IVF conceptions from 30% to about 10% coincident with the SET mandate, with a further decline to 5% from 2004, after which it stabilizes. At 5% this is still well above the share of twins among unassisted births (also displayed in the Figure) of about 1.62%, consistent with close to 30% of post-reform IVF births proceeding with DET and with the biological tendency for twin births to increase with parental age and socio-economic status (Bhalotra and Clarke, 2019).

Maternal and child outcomes improve sharply after the SET mandate. As a prelude to the analysis, we show unconditional outcome data plots in Appendix Figures B3, B4, B5 and B6. Most of our outcome indicators for child and maternal health and for maternal earnings show a sharp improvement after the 2003 reform. We do not dwell upon the single difference shown in these plots but instead study the double difference event study plots. We now describe the strategy for generating these.

[24] The birth registers do not identify SET vs DET within IVF births, but the share of SET births is available in aggregate data published by the National Board of Health.

4 Empirical strategy

We use a panel event-study design to draw causal inference over how the 2003 IVF-SET reform, which created quasi-experimental variation in the risk of twin birth, impacted fertility, child and maternal health, sickness benefits, and mother's and father's earnings. The reform occurs nationwide, and a single difference is vulnerable to capturing the influence of omitted trends that change in or after 2003. In order to control for aggregate trends, we difference the outcomes of women using IVF with respect to the outcomes of women not using IVF. The estimated equation is:

$$Y_{it} = \alpha + \sum_{k \in \ell} \gamma_k (IVF_i \times \mathbb{I}\{Year_t = SET + k\}) + \beta IVF_i + \boldsymbol{X}_{it}\delta + \alpha_c + \pi_t + \nu_{it}, \quad (1)$$

where the dependent variable Y_{it} refers to an outcome for birth i in year t, and IVF_i refers to the IVF status of each birth (1 if IVF was used, and 0 otherwise). We interact the IVF indicator with four leads and lags on either side of the reform, $\ell = \{-4, -3, -2, 0, \ldots, 4\}$. The year before the reform, 2002, is omitted as a base category. 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. This is investigated by testing the equality of the lead (pre-SET) coefficients. The lagged (post-SET) coefficients capture the dynamic effects of the reform.[25]

Year$\times$month of conception fixed effects π_t control flexibly for all relevant time varying unobservables. County of birth fixed effects α_c capture time-invariant geographical variation in the outcomes. The control variables $\boldsymbol{X}$ are fixed effects for the following characteristics of both mothers and fathers: age at birth, education level, whether native (defined as born in Sweden); the pregnancy order (birth order) of the mother; and the gender of the child. To account for their potential impact on twin birth (Bhalotra and Clarke, 2019), we also control for whether the mother smoked in the first trimester of pregnancy and her BMI before pregnancy. Standard errors are clustered by mother.

We complement the event study plots with estimates of the single coefficient two way fixed effects equation:

$$Y_{it} = \alpha + \beta_1 (PostSET \times IVF)_{it} + \beta_2 IVF_i + \boldsymbol{X}_{it}\delta + \alpha_c + \pi_t + \varepsilon_{it}, \quad (2)$$

where all variables are defined as in equation 1, but we now cumulate all of the lag coefficients into a single indicator variable, $PostSET$ defined as 1 for all births conceived after January 1 2003, and 0 for those conceived before. The parameter of interest is β_1, capturing the average change in outcomes for IVF births relative to non-IVF births after 1 January 2003. For the health and fertility outcomes, we will report the extent to which SET led to a convergence of IVF birth outcomes towards non-IVF birth outcomes, defined as (β_1/β_2). In the case of earnings, IVF users

[25]In this specification, year refers to conception year. The post-SET coefficients describe the change in the outcome for children conceived in 2003, 2004,…, 2007. For each conception year, earnings are defined as an average over the nine years after birth. We similarly take an average over nine years for child hospitalization.

start out with an advantage so we provide the more conventional normalization of the coefficient on the pre-reform mean. As there was a single adoption date which holds nationwide, we need not be concerned about problems in weighting in two-way fixed effect models with heterogeneous treatment effects that have been shown to potentially create bias in staggered adoption designs (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020; Sun and Abraham, 2020). Heterogeneous effects will be reduced to mean pre-SET versus post-SET changes between IVF and non-IVF mothers in equation 2.

Counterfactual trends. We plot event study estimates of equation 1 with confidence intervals to allow the reader to assess if outcomes for women who did and did not use IVF exhibit differential pre-trends. Even if we cannot reject parallel pre-trends, this may be because we are under-powered to do so (Roth, 2021). For this reason, we implement recent partial identification methods that estimate upper and lower bounds on the estimates after relaxing the assumption of parallel pre-trends (Rambachan and Roth, 2020). To do this, we need to make assumptions on the degree of violation of parallel trends allowed. We estimate pre-trends between years -4 to -2 (omitting year -1) and project these forward as post-trends, allowing them to vary by some amount M between each subsequent period. We plot a series of estimated bounds corresponding to a range of values of M.[26]

We know that there is selection into IVF, for example, women using IVF are older and more educated than other women. The DiD estimator only requires trends, not levels to be parallel, but level differences may lead one to worry that the control group (women not using IVF to conceive) does not provide a good counterfactual for the treated group (women using IVF). In the baseline specification we control for relevant parental characteristics using flexible specifications. We estimate an additional model that controls for parental characteristics interacted with a linear trend. This accounts for differences in outcome trends that derive from compositional differences, for example, it allows that fertility or earnings evolve differently for more vs less educated women.

The SET reform as a policy experiment Equations 1 and 2 are reduced form representations of a system in which the SET mandate instruments the twin birth rate, with the drop in twinning driving the other outcomes. We provide estimates of the first stage, which demonstrate instrument relevance (power). The exclusion restriction that the SET reform impacts the outcomes only through twinning is plausible given that SET involved a switch from two to one embryo with no concurrent changes in IVF treatment procedures with respect to medication, technique or equipment (Saldeen and Sundström, 2005). As explained earlier, the SET mandate permitted exceptions, allowing women with lower chances of pregnancy success, including women over the age of 38, to elect DET. The birth register has an individual tag for IVF but it does not reveal which IVF-users

[26]These methods at baseline allow for any prevailing differential trends between IVF and non-IVF outcomes. The values M then allow for *additional* movements beyond those implied in any prevailing trends. Values of M are naturally outcome dependent, for example they are allowed to be up to 1% in the case of twinning, 10 grams in the case of birth weight, 1,000 SEK in the case of income, and so forth. The presentation of bounds over a range of values of M allows varying priors related to these values to be considered, and as such, we present a series of bounds gridding over M.

used SET rather than DET. As a result, our estimates are intent to treat (ITT) estimates.

That the SET mandate was exogenously driven by scientific research (on pregnancy success rates in trials comparing SET with DET) and that its abrupt introduction disrupted reproductive health outcomes is clear. However, unconditional plots of trends in the outcomes we analyse (Figures B3-B6) reveal that many outcomes began to move in the direction stimulated by SET from 2001. This is consistent with the fact that medical professionals had the evidence before the mandate and that, under their guidance, there was an elective increase in SET before the mandate. Ignoring this in the baseline specification makes our results conservative (as the jump at the date of the reform is smaller). However, we show that our results are robust to dropping the two years of gradual increase so that identification comes from a sharper discontinuity in the share of twin births to IVF users. This specification check also takes care of the fact that the region of Skåne implemented SET in 2001, ahead of the rest of the country. Allowing this window around the SET reform also addresses measurement error in the date of conception, which is used to assign births to the post-SET (treated) vs the pre-SET (control) cohorts.

Endogenous changes in sample composition We have argued that the SET mandate provides a valid policy experiment. Accordingly, our estimates will identify the causal impact of the mandate. If we further wish to interpret our estimates as identifying impacts of replacing twin birth with singleton birth – then we need to examine endogenous changes in sample composition. These can arise on account of a change in the sorts of women that undertake IVF after SET, or on account of changes in IVF success rates. We showed in section 2, using published aggregates, that the IVF success rate did not change. This is in line with Lundborg et al. (2017) who demonstrate that IVF success rates in Denmark are idiosyncratic, being uncorrelated with a rich set of observable characteristics of the woman. We now consider how we assess the other concerns.

First, note that if women with more favourable maternal characteristics were more likely to select into IVF after the SET mandate then this would lead to higher twinning alongside better health and earnings (Bhalotra and Clarke, 2019), while the SET reform led to lower twinning alongside better health and earnings. Second, we consider whether the share of women seeking IVF changed discontinuously at the date of the SET reform. If it did, it is plausible that women at the margin are different and hence that there is a compositional shift. Figure A1b shows no structural break in the trend in women using IVF coincident with SET. Third, we investigate endogenous changes in composition by regressing each characteristic of women (and their partners) on IVF and PostSET$\times$IVF in what is effectively a test of balance. A statistically significant coefficient on PostSET$\times$IVF is indicative of selection, and we will see that there is limited evidence of this.

In any case, as discussed, we also account for any differences by controlling flexibly for characteristics of the mother and father in the model and for trends in these characteristics. If there were endogenous selection into the IVF sample on characteristics then the controls for characteristics would be potentially endogenous. We therefore show results without these controls, and the results are broadly similar.

As our consideration of selection into IVF can only be based on observable characteristics of parents and families, in a further specification check we allow for unobserved mother-level unobserved heterogeneity by including mother fixed effects. As mother fixed effects in a model with a binary interaction (of IVF with SET) can result in undesired comparisons, we create an explicitly matched sample that allows us to compare mothers exposed vs unexposed to the SET reform, so that we isolate impacts of the SET reform on *within* mother variation. A discussion of the issues that arise in the standard model, and of our matched sample and results is in Appendix C.

In another approach to accounting for selection on unobservables, we conduct a partial identification exercise. We implement the procedure in Oster (2019), who extends the estimator suggested by Altonji et al. (2005). The approach involves considering how likely it is that our results can be explained by selection into IVF based on unobservables. We estimate bounds on our estimates on the assumption of equal selection on observables and unobservables, and under a range of assumptions about the degree to which unobservable factors could explain the outcomes of interest.

Multiple hypothesis testing We have many indicators of the outcomes. We are thus faced with a problem of multiple-inference and risk over-rejecting null hypotheses (i.e. an inflated rate of Type I errors). We address this issue using two different approaches. First, we adjust all p-values by controlling for the family-wise error rate (the proportion of Type I errors committed among *any* of the outcome variables considered) among all variables examined in the paper using Holm's step-down procedure.[27] This method has the advantage of greater power compared to single step approaches such as Bonferroni. It is more demanding than false-discovery rate corrections which set the error rate based on the proportion of Type I errors in all significant findings. Second, we create summary indices for child health and maternal health which aggregate multiple measures of morbidity, thus decreasing the number of hypotheses tested. The impacts on individual components of morbidity indexes are displayed in Tables A9 (children) and A10 (mothers). The indexes take the value of 1 if at least one morbidity occurred, and 0 otherwise.

Outcome distribution and heterogeneity by characteristics For key continuous outcomes we estimate and display impacts of the SET reform across the outcome distribution. To do this we estimate equation 2 several times, defining the outcome as a binary variable indicating that the outcome exceeds a range of specific points along the distribution of outcomes, following Rossin-Slater (2013). We also investigate heterogeneity in impacts of the reform by mother characteristics including parity, education, age and body mass index (BMI). Levels of each of these categories, as well as rates of IVF birth in each group, are documented in Table A8.

[27] This is a 'step-down' procedure in that the most significant p-value is multiplied by K, the second most significant p-value is multiplied by $K-1$, the third most significant by $K-2$, and so forth, where K is the total number of hypotheses to be tested. This procedure could be further refined using step-down methods which additionally consider correlation between outcomes of interest, for example methods of Romano and Wolf (2005). Here we use Holm's method given its ease of application across various data sources, but our corrections should thus be viewed as conservative.

5 Results: Twin birth, fertility, child and maternal health

Event study estimates of the impact of the SET reform on IVF outcomes are in Figures 2-5, and the corresponding single coefficient estimates are in Tables 1 to 4. Estimate showing how the outcome distribution is shifted with SET, for key continuous outcomes, are in Figure 6, and Figure 7 describes birth spacing, or dynamic fertility impacts. For parsimony, we provide additional results for selected key outcomes in the main text, and for all other outcomes in the Appendix. Rambachan and Roth (2020) bounds on the dynamic effects are in Figures 8 and Appendix Figures A2-A3. A sequence of additional robustness check is displayed in Figure 9 and Appendix Figures A4-A5. Heterogeneity in treatment effects is displayed in Figures 10 and Appendix Figures A6-A7.

In all results tables, the coefficient in the first row displays the impact of the SET reform, identified from a double difference that engages pre- vs post-SET variation in IVF relative to non-IVF birth outcomes. The coefficient in the second row provides the main effect of IVF on the outcome. For fertility and health outcomes, we consistently provide the ratio of these coefficients as the scaled impact of the SET reform, or the extent to which the passage of SET narrowed the baseline gap between IVF and non-IVF births in the outcome. For brevity, we will sometimes refer to this as "the gap". For earnings, because the earnings of women using IVF are higher than of other women before the reform, we scale these estimates, as is standard, by their pre-reform mean (i.e by the baseline IVF mean). Every table reports the p-value attached to the double difference coefficient unadjusted as well as adjusted for multiple hypothesis testing (henceforth MHT). Most coefficients that are significant unadjusted remain significant after the MHT correction. We only explicitly discuss the MHT correction in cases where the adjustment results in significance being lost.

5.1 Twin birth rate

See Table 1. IVF pregnancies in the pre-reform period were 17.5 percentage points (pp) more likely to be a twin pregnancy. The SET reform leads to a decline in the risk of twin birth of 12.3 pp. Thus the SET reform narrowed the gap in twinning between IVF and non-IVF births by about 70%. The data do not identify at the individual level which women used SET and which were exempt and allowed to use DET, though we know from aggregated figures provided by the government that about 70% of post-reform IVF births used SET, versus 10% of pre-reform births. That twinning fell by 70% while rates of SET increased by only 60% is consistent with the fact that in the post-reform period, DET procedures were targeted at women with a low probability of twinning.

The event study plot is in Figure 2a. The dynamic impacts of the reform reveal that all of the decline occurred within a year after reform, after which it stabilized, consistent with the reform being a mandate. Now consider the pre-reform coefficients. The difference in the share of twin births between IVF and non-IVF women fluctuates in the pre-reform period, showing no trend until it declines steadily from 2002 (-1 in event time) to 2004 (+1 in event time). As discussed earlier, this is consistent with medical professionals having learnt of the potential for single embryo transfers to preserve pregnancy success rates before the National Board of Health mandated SET. Specification checks discussed in the next section establish that our results are robust to dropping the transition years.[28]

5.2 Fertility

There are two reasons that women may want to avert a twin birth. First, they may desire one child rather than two children. Second, they may desire two (or more) but prefer them spaced out. In the latter case, women using IVF are more likely to proceed to have another birth event after SET relative to before. Understanding the relative weight of these preferences among IVF users is relevant to understanding the welfare impacts of mandating SET. We model the number of births and the probability of at least one further birth (modelled as the inverse, which is no further birth), see Table 1, columns 2 and 3.

Prior to the SET reform, women conceiving after IVF were 10 percentage points more likely to cease fertility (row 2, column 3), perhaps principally because they were more likely to have had twins. After SET, there is a 7.2 pp increase in the chances of IVF users continuing fertility (row 1). Thus, after SET is mandated, a higher share of IVF births is followed by a subsequent birth. Thus SET closes about 70% of the pre-reform gap.[29] Column 2 shows that, before SET, IVF users had 0.11 fewer births in the nine years following the index birth than women who do not use IVF (row 2). After SET, this gap narrows as IVF users have 0.07 more births than before. The event study plots in Figures 2b-2c show significant and persistent post-reform changes and the pre-reform differences show no trend.

Clearly women at first parity are more likely to continue fertility than women who already have children. We therefore repeated this exercise distinguishing women having their first birth from the rest, see Figure 10, panel (b), as well as Appendix Figure A7. As shown in Table A5, 74% of all IVF births are first births, while 44% of all non-IVF births are first births. The post-SET increase in the chances of a second birth is entirely driven by women for whom the index IVF birth is a first birth: in this sample SET results in an increase in the number of births by 0.11, and a decrease in the chances of no future birth by 10 pp. Among women whose index birth was second or higher, both of these coefficients are essentially zero.

Figure 7 considers the timing of these additional births. We estimate specification 2, but rather than modelling total future fertility, we estimate a series of separate regressions where the outcome

[28] We also observe in Figure 2a that the pre-2002 coefficients, while more or less flat, do not lie on the zero line but, instead, are shifted up. This is mechanically because of the slight rollout of SET pre-reform, which acts to push the second and earlier leads above the first lead, which is the omitted base. It is for the same reason that the pre-reform coefficients in Figures 2b (discussed next) are shifted to lie below the zero line, and our remarks here apply to all event studies discussed. Dropping the transition years of 2001-2002 would result in the pre-coefficients resting on the zero line. In any case, note that the level of the pre-reform coefficients does not challenge the pretrends assumption.

[29] The unconditional baseline means displayed in the Table show that, averaging over parity, before passage of SET, 46% of IVF users and 53% of women having a natural birth continue to have another child.

is an indicator for a woman having a birth $k \in \{1, \ldots, 7\}$ years post-birth, and the independent variable of interest is the same $PostSET \times IVF$ terms. We observe that additional marginal births following SET generally occur two or three years after the index birth.

5.3 Child health outcomes

See Table 2. Neonatal outcomes for children born of IVF show a dramatic improvement across a range of indicators after the SET mandate. IVF newborns are 194.7 grams heavier at birth, which corresponds to 61% of the baseline IVF deficit relative to unassisted births. They are less likely to be born prematurely, having 0.61 weeks longer gestation, which is 56% of the baseline gap. The probability of a low APGAR (less than 7 in a range of 0-10) is lower after SET, narrowing the gap by as much as 75%. The APGAR is a test done 1 and 5 minutes after birth that checks a baby's heart rate, muscle tone, and other signs to see if extra medical care or emergency care is needed. We investigated numerous other indicators of health at birth including birth length, head circumference and breech presentation, SET-led improvements in which narrowed the IVF–non-IVF differential by close to 50%, these results are available in Bhalotra et al. (2019), where we also report no discernible change in the probability that the IVF birth is male and the probability of fetal malformation.

There is a decline of 3.1 pp in the probability that IVF births suffer at least one severe neonatal morbidity outcome, closing the gap by 56%. The 17 morbidities included in the index are detailed in Table A2 and Table A9 shows morbidity-specific results. After using the MHT adjustment across the 17 outcomes, five show a significant reduction. These are extremely low birth weight, extremely preterm birth, hypoglycemia, hypothermia and meconium aspiration. In line with reduced morbidities at birth, post-SET IVF births spend 1.8 fewer nights in hospital after birth, narrowing the gap by 62%. They are also less likely to return to hospital later on, between the ages of 1 and 8 – although this coefficient is not statistically significant, we note that there is a significant gap (i.e. IVF births spend significantly more nights in hospital after the infant period) and that SET narrows this by 38%. Child mortality rates in Sweden are low and there are no statistically discernible impacts of SET on infant or under-5 mortality, but, again, there is a significant baseline gap and, after the SET mandate, while imprecise, there is a sizeable reduction in these gaps of 42 and 76 percent respectively. On account of imperfect compliance (generated by allowed exceptions to SET), together with the fact that IVF singletons have worse health than singletons born following an unassisted conception (shown in preceding section), we do not expect absolute convergence of IVF to non-IVF outcomes.

The event study plots in Figure 3 provide a vivid depiction of the results, additionally demonstrating persistence of the post-reform effects and showing no significant differences in pretrends between IVF and non-IVF births. The impacts of SET tend to be larger among first births, but the differences are often not significant, see Figures 10 and A6. The significance of the improvement in the APGAR score is not robust to MHT, but the other results are.

Distribution of effects For three of the most commonly used indicators of birth outcomes, birth weight, gestational age and the APGAR score, we estimate distributional impacts, see Figure 6. We observe significant impacts of SET across most of the distribution for these outcomes. For birth weight, the largest marginal impact is in the middle at around 3,000 grams, while for gestational age and APGAR the largest impacts are in the upper regions, around 37 weeks and for scores above 6 respectively. In these plots, we have overlaid the baseline (pre-SET) distribution of the outcome in the IVF group, which is a measure of reform exposure. For all of the birth outcomes considered, we see larger than proportional SET-treatment impacts in the lower tail and smaller than proportional impacts in the upper tail. This makes sense because the SET reform reduced the risk of twin birth and twins have lower birth weights, gestational ages, and APGAR scores than singletons.

We also studied marginal effects at thresholds that are often used in the targeting of medical resources (Almond et al., 2010; Bharadwaj et al., 2013), which are presented at specific distributional points in Figure 6a. We estimate that, following the SET reform, the likelihood of IVF babies being born with a weight below 1,500 grams (very low birth weight) fell by 1.2 pp, and the likelihood of being born with a weight below 2500 grams (low birth weight) by 7.8 pp. Scaling by pre-reform differences between IVF and non-IVF babies, the proportional impacts are both about 60% (full calculations are documented in Bhalotra et al. (2019)). The probability of preterm delivery before weeks 28, 32 and 37 decreased by 0.5 (63%), 1.3 (52%) and 8.3 (53%) pp. respectively. Very similar results are found for first-time mothers.[30]

5.4 Maternal health outcomes

See Table 3. Using an index of maternal morbidities around childbirth (defined in Table A1), we estimate that SET results in 1.1 pp decrease in the chances that mothers suffer at least one of a set of morbidities, corresponding to a 20% narrowing of the gap between IVF and non-IVF mothers. The baseline gap is large and the fact that 80% of the excess morbidity women incur under IVF remains after SET flags risks to women associated with IVF over and above the risk of having a multiple birth.

The 18 morbidities included in the index are detailed in Table A1 and Table A10 shows morbidity-specific results. Pre-eclampsia and an indicator for multiple birth complications show a decline after SET, but these do not survive adjustment for testing 18 hypotheses. The number of nights mothers with an IVF birth spend in hospital following childbirth declined by 0.63 on a base of 5.2, constituting a narrowing of the gap by a considerable 63%. The probability that the birth is an emergency C-section declined by 1.5 pp, narrowing the gap by 35%. These outcomes are not independent, as women who have a C-section birth are more likely to spend longer time in hospital, and more likely to suffer maternal morbidities. The result for morbidities and C-sections is not

[30]When we say that SET makes it around 8% more likely that a birth exceeds 2,500 grams this can be because it shifted births from 2,450 to 2,500, or from 1,000 to 2,500. We can only say that the distribution moves to the right without saying from what baseline.

robust to MHT, the result for hospital nights is. In all cases, results are larger for first time mothers than mothers having a higher order birth (Figure A7).

Another indicator of impacts of SET on maternal health that captures any morbidities that limit attendance at work is sickness benefit, defined in the Data section. We find a significant reduction of 665 SEK in receipt of sickness benefits after SET among women using IVF, which narrows the baseline gap by 40%, see columns 2 and 4 of Table 4. The result for fathers is in the same direction but imprecise. Significance of the decrease in women's sickness payment receipts is robust to MHT. Figure 4 provides event study plots for the outcomes in Table 3 and Figure 5 for sickness benefits. They highlight that this set of results is estimated with relatively low precision. Among mothers having their first birth, the impact of SET on maternal morbidity and emergency C-sections is larger and, also, tends to be estimated with precision (Figure A7), but the impact on sickness pay is similar for first and second births (and null for higher order births above birth 2, see Figure 10).

6 Results: Parental earnings and the child penalty

In this section we describe impacts of the SET reform on the earnings of mothers and fathers.

6.1 Mother's earnings after birth

See Table 4. While child and maternal health at baseline (pre-SET) are consistently worse for IVF births, mother's earnings are higher among women who have IVF-assisted births. This is a reflection of the stylized fact that there is positive selection of women on socioeconomic status and age into IVF treatments. For each post-SET birth cohort, earnings are defined as an average over the 9 years after birth. Averaging over the four post-SET cohorts, we identify an increase of about 5.6% following the SET mandate. Thus SET widens the baseline gap in favour of IVF-using women (column 1).[31] The event study plot (Figure 5a) shows that the increase in women's earnings persists through the four cohorts born after the SET reform. Significance of the increase in women's earnings is robust to MHT.

Our finding that mothers giving birth with IVF suffer a smaller child penalty in earnings after SET is consistent with an indirect impact that derives from improved maternal and newborn health, and a direct impact of lower rates of twin birth. We have already documented that adverse health impacts (on mothers and children) are most clearly evident in the year of birth. In the next section we investigate earnings dynamics following twin vs singleton birth and identify an excess child penalty from twin birth which, similarly, is evident only for a short period (two years) after birth, after which there is convergence of the birth penalty from twins towards that from singletons, with

[31]The 5.6% is calculated as the SET driven increase of 10,872 SEK in row 1 divided by the baseline IVF mean of 192,983 in row 5. The scaled impact, for consistency with the other tables, is the ratio of the coefficients in rows 1 and 2 but since, in this case, the baseline gap favours IVF users, the scaled impact is a less meaningful statistic.

the long run penalty tending to slightly favour women who start out with twins.

Distribution Figure 6d shows that the SET reform acted to shift earnings towards the middle of the distribution. The likelihood of having a salary of about 250,000 SEK increases by nearly 4 pp. We estimate smaller, insignificant increases in the lower half of the distribution, though the point estimates suggest consistent positive shifts in the likelihood of having a mean salary exceeding all distributional mass points considered.[32] When considering baseline distributions of earnings, unlike the case of health outcomes where we observed larger than proportional impacts lower in the earnings distribution, in the case of earnings we observe that impacts are larger than proportional at relatively higher points of the earnings distribution. Thus, even though impacts are observed across the earnings distribution, these results suggest that relatively higher earning women gain relatively more from the reform.

Extensive vs intensive margin changes The results discussed so far include cases where earnings are zero. We separated these cases to illuminate impacts on women's employment as distinct from impacts on their earnings (Appendix Table A11). The SET mandate led to a significant increase in women's earnings at both the extensive and intensive margins. The increase at the extensive margin is small, at 0.5 pp, consistent with the baseline share of women with a non-zero wage being 97%. The intensive margin increase is therefore very similar to the total increase shown in Table 4.

First vs higher order births Estimates for women having their first vs higher order birth are in Figure 10. In contrast to the results for maternal health, SET has a larger impact on earnings for women for whom the index IVF birth is second or higher order – in fact the coefficient is twice as large, roughly 20,000 SEK rather than 10,000 SEK, though confidence intervals overlap in each case.

6.2 Father's earnings after birth

Sweden pioneered parental leave policies starting in 1975. Parents in Sweden are now entitled to 480 days of paid parental leave, with 180 extra days for twin births.[33] In 2019, fathers took, on average, 30 per cent of all paid parental leave. Thus, in principle, fathers' earnings could suffer from having a child. On the other hand, fathers may catch up after paternity leave, working harder, as suggested by Lundberg and Rose (2002) for example. We find no impact of SET on the earnings of men who father children born with IVF (column 2). Observe that the baseline gap for men, as for women, favours IVF-users (rows 5 and 6) . Event studies for mother and father earnings are in Figure 5, showing a clear upward shift for women's earnings post-SET, and a flat profile pre- and post-reform for fathers.

[32]Since earnings are in SEK and not in logarithms, equal proportional increases will show as larger absolute increases.

[33]Although the parental leave program is individual, allowing 240 days for each parent, all days above 90 can be transferred to the other parent. The number of non-transferable days increased from 30 in 1995 to 60 in 2002 and to 90 in 2016.

Relative income of women Since SET resulted in an increase in women's and not men's earnings in the IVF sample, it follows that it was associated with an increase in the relative earnings of women in the aggregate. Motivated by the common assumption that women's bargaining power within the family is a function of her income relative to that of her partner, we used individual linked data in the register to create the relative female wage at the household level.We find that, on average within IVF couples with an IVF birth, there is a 6.2% increase in the relative earnings of women after SET (Appendix Table A11). Significance of this result (as well as the extensive margin earnings coefficient for women) does not survive the MHT adjustment, though we note that multiple hypothesis adjustment here is quite demanding, being based on a FWER correction which conserves size at the cost of power.

7 Additional Results

In this section we provide robustness tests on the full sample results, estimates of heterogeneity in impacts of the reform by characteristics of the mothers, and a sketch of our approach to a cost-benefit analysis of the SET reform.

7.1 Robustness checks

In the empirical strategy laid out in Section 4 we set out the identifying assumptions, noted challenges to identification and explained how we address these. In this section we provide the results of specification checks and extensions, referring the reader to Section 4 for the motivating principles.

Differential trends for IVF vs non-IVF outcomes The event study plots discussed in the previous section, in general, show no evidence of differential pre-trends. However, fluctuations in the pre-reform coefficients are consistent with outcome trends changing a year before the reform. This is in line with new scientific evidence being disseminated in advance of the mandate, leading medical professionals to advocate elective SET from 2001. To address this we re-estimate SET impacts dropping the years 2001-2002, and the coefficient of interest is close to identical, see Figures 9, A4 and A5. This specification checks also confirms that our estimates are not sensitive to measurement error in the date of conception, which we estimate using information on birth date.[34]

We nevertheless investigate sensitivity to adjusting for trends. First, we control for every par-

[34]The estimates are also robust to removing the region of Skåne which mandated SET in 2001. These results are not shown because, in principle, people could migrate to Skåne for the procedure and give birth in their home county, they are however available in an earlier working paper version of the paper, (Bhalotra et al., 2019). We also note there that in 2005 Sweden started to offer same-sex couples publicly funded access to fertility treatments including IVF. Same-sex couples tend to have higher socioeconomic status (Ahmed et al., 2011b,a) but their children tend to have worse birth outcomes, at least as observed in lower birth weight (Aldén et al., 2017). To account for this legislative change we restricted the sample to conceptions occurring during 1998-2004. The estimates are similar but for parsimony, are not displayed as the number of children born to lesbian parents during 1995-2010 is only 750.

ent characteristic interacted with a linear trend. This allows that the outcomes (e.g. birth weight, or women's earnings) evolved differently over time for women with different characteristics (e.g. education or age). If they did, and if IVF-users had different characteristics, then this could manifest as a violation of the pre-trends assumption. Estimates conditional on trends in characteristics are not significantly different from the baseline estimates, as documented in the plots labelled "Trends in Parent" in Figures 9, A4 and A5.

The same Figures display estimates of a more demanding specification that additionally allows the trends in parental characteristics to break in 2003, the date of the SET reform. This adjusts more flexibly for selection into IVF, and also adjusts for selection into SET. The magnitude of the treatment effect now moves a bit but, in general, the baseline results continue to hold.[35] The stability of the results to these specification checks is consistent with our finding, reported in the previous section, that impacts of the SET reform on the analysed outcomes are, in general, evident across parents with different age, education and health characteristics.

To allow that the event studies are underpowered to detect pre-trends, and additionally to allow for a more flexible class of parallel trend assumptions where counterfactual trends can *further* diverge from linearity between subsequent periods, we estimate bounds on the dynamic post-SET coefficients using the honest DiD estimator of Rambachan and Roth (2020). See Figure 8 for key outcomes and A2-A3 for all other outcomes. The baseline estimates largely stand up to this specification check.[36]

Consider for example rates of twinning, documented in Figure 8a. Pre-event lead coefficients from -5 to -2 suggest a very moderate downward trend if a line is traced through point estimates in these periods (including a baseline 0 in year -2, as evident also in Figure B3a). We can trace this forward as the counterfactual trend in place of a parallel trend strictly at the zero line, which suggests that rates of twinning may have slightly decreased among IVF women compared to their non-IVF counterparts, even in the absence of SET. Considering this slight decline as the counterfactual trend thus shifts confidence intervals in the direction of the zero line, though they are still located significantly below 0. Additionally, we allow further divergence from linearity by the values indicated as M. For example, in the most demanding case, we allow rates of twinning to further diverge by as much as 1 percentage point per year when considering counterfactual trends, generating bounds based on most the extreme lower and upper confidence intervals encountered. In each of lags 0, 1, 2 and 3, the bounds are still informative under these quite demanding assumptions, becoming insignificant only at lag 4 given the accumulation of uncertainty over time. Similar such interpretations can be made for all outcomes and bounds plotted in Figures 8, A2 and A3.

[35]Inclusion of trends that are allowed to break in the year of SET tends to reduce the impact of SET, except for infant mortality and sickness benefits when it increases it but, in general, the new estimates are not significantly different from the benchmark.

[36]Among notable changes are that the APGAR score no longer shows a significant improvement, but infant mortality shows a significant decline now having been insignificant earlier. These are two of several outcomes.

If one is concerned about omitted trends biasing estimates, it is often useful to be able to identify a placebo experiment. Since we investigate impacts of a mandate and there were documented exceptions to the mandate, we are able to leverage this. We estimate impacts for women over the age of 38 as they were exempt from the mandate. See Figure 10 and Appendix Figures A6-A7 which show that, consistent with this, in general, we see either muted effects or no effects of SET in this group. This is not the ideal placebo because older women needed to electively opt-out of SET (and indeed, we observe a small reduction of twinning even in this group), and because older women may have different outcomes for biological reasons.

Endogenous selection into IVF The share of IVF births in all births has increased secularly since the 1990s, tracking changes in technology, costs, and availability of IVF, and there is no evidence that SET prompted either a higher or lower share of women to opt for IVF, see Figure 1a. However, if SET prompted a change in the composition of IVF-using mothers this would modify the extent to which we can interpret our estimates as impacts of replacing twin birth with singleton birth. Note that our estimates would, irrespective, identify impacts of the SET mandate, a clearly policy relevant parameter.

We investigate this directly by regressing each of a host of available mother and father characteristics on the treatment term, in the spirit of a test of balance, see Table A12. Of the twelve characteristics, only one has a MHT adjusted p-value that indicates imbalance: after SET, the father of the child of an IVF birth is 1.9% more likely than before SET to be native-born.[37] In any case, these differences are small, and we condition flexibly upon all available characteristics. We assess sensitivity of the estimates to this by producing estimates in which we drop all controls for parental characteristics. The coefficients hardly move, see Figure 9 for key outcomes, Appendix Figures A4-A5 for the rest.

Finding limited selection on a range of observables may lead us to expect limited selection on unobservables. We investigate this formally below. We first investigate selection on unobservables fixed at the mother-level by estimating a specification with mother fixed effects. In practice, when the independent variable of interest is an interaction term the standard approach does not provide meaningful estimates, essentially because the identifying variation includes comparisons that are not sensible. This is explained in Appendix C, where we identify the subset of comparisons that is sensible. Pooling these, we estimate the mother fixed effects specification on a sample restricted to women with at least two pregnancies (about 50% of all IVF mothers). Table C1 shows that the SET mandate leads to significant increases in fertility, birth weight, infant survival and earnings in this selected subsample, in line with the benchmark estimates.[38]

[37] The *unadjusted* p-values additionally point to an increase of about 4% in the share of both mothers and fathers with tertiary education, which may reflect an underlying trend in education together with the fact that IVF-users are consistently more educated than other women.

[38] We find no significant impacts on the other outcomes. In principle this could be because the other outcomes are driven by endogenous selection of women into IVF (unlikely given the balance tests and robustness to adding vs dropping controls for characteristics, as well as the trend tests discussed next), or by the sample being selected by construction (unlikely given similarity of the estimates for the outcomes for

Table 5 investigates whether our estimates can be explained by selection on unobservables, now allowing that these vary across mothers and time. Following Oster (2019) we estimate how much more important than observables the role of unobservables would need to be for the estimates to be driven down to zero. Oster (2019) argues that the coefficient stability assumptions formalized in Altonji et al. (2005) require some conception of the proportion of all outcome variance that could be explained if all relevant unobservable factors were included. We present bounds under two assumptions. First, following Oster (2019) we assume that unobservables are as important as observables, and that the addition of all relevant unobservables would be sufficient to explain 1.3 times the R^2 in the model with observable controls. Second, we obtain "extreme bounds", assuming a maximum R^2 of 1.

Panel A of Table 5 presents the estimates for child outcomes, and Panel B for mother outcomes. Consider the results in column 1 of Panel A. Unconditional on observable controls, the estimated impact of SET is that birth weight increases by 200 grams. Conditional upon observable controls, this impact falls to 194 grams. When moving from the unconditional to conditional models, the R^2 increases from less than 0.01 to 0.071. At the base of the Panel we first present values for δ following the terminology of Oster (2019), which captures how important selection on unobservables would need to be relative to selection on observables to drive the estimate of SET to 0. Now assuming that the maximum R^2 is 1.3 times the R-squared in the model with controls, we estimate that unobservables would need to be nearly 100 times more important than observables for the true effect to be 0. This large degree of selection seems highly unlikely. We then present bounds on the estimates assuming that unobservables are as important as observables and $R_{max} = 1.3\times$ the R^2 from the controlled model. In the case of birth weight, where the addition of observable controls does little to move the estimate on SET, we see that these bounds are quite tight, ranging from 192.7 grams to 194.7 grams. Finally, we consider the much more extreme case where we assume that unobservable controls could completely explain all variation in birth weight. In the case of birth weight, even under these extreme bounds the lower bound only falls to 111.36, implying considerable robustness to unobservable selection into IVF after SET.

Looking across outcomes, a number of broad interesting patterns emerge. First, for all outcomes of interest, unobservables would need to be considerably more important than observables for the true results to be zero. Among significant effects, these range from 11.5 in the case of wage earnings, to 126 in the case of sickness benefits. Given that a criteria of 1 is often employed, these values suggest that it is unlikely that results owe to selection on unobservables. Secondly, the bounds following Oster (2019) are uniformly informative. In many cases where the significantly more demanding extreme bounds are presented, significant values would still be estimated, with the impacts of SET being 111 in the case of birth weight, a 0.7% reduction in the case severe neonatal morbidity, a 2.5pp reduction in the case of twinning, or a 4800 SEK increase in wages. Where these extreme bounds are not informative is principally in the case of health outcomes where the

which significance persists with mother fixed effects), or because we are underpowered to detect effects. The last seems most likely– the mother FE estimates in Table C1 are estimated on barely 5,000 observations, while the benchmark sample contains about 900,000 observations.

explanatory power of models is generally low, and in which case assuming that unobservables can fully explain outcomes is a particularly demanding (and unreasonable) assumption. The sequence of tests of selection on mother-level observables and unobservables together indicates that our estimates of the causal impact of SET also provide a good approximation to the causal impact of shifting from twin to singleton birth.

Misclassification We conclude with a discussion of possible misclassification. Comparing the Medical Birth Registry with national IVF data, it seems that it correctly identifies between 70 and 90 percent of all IVF births. Thus 10 to 30 percent of IVF births may be incorrectly recorded as non-IVF births. This is a second order concern for two reasons. First it will contaminate the control group and lead to under-estimation of the impacts of SET. Second, the size of the treated group (IVF users) is so much smaller than the size of the control group (mothers who do not use IVF) that, even if the impact of the reform is very large, it is unlikely that the 30% of mis-classified IVF births will impact averages in the control group in any substantive way. To see this, consider that the number of observed IVF births in the Medical Birth Registry is 22,183, and the number of non-IVF births is 932,822. Inflating the number of IVF births from 70 to 100% implies that 9,507 IVF births are incorrectly classified as non-IVF births. This is only slightly more than 1% of non-IVF births. We provide additional discussion, as well as a calculation of the (small) magnitude of any expected attenuation for the worst case of 30% mis-classified in Appendix D.

7.2 Heterogeneity in impacts of the SET reform

We investigated heterogeneity by categories of mother's parity (birth order), age at birth, education and BMI. The shares of women in each of these groups who use IVF are provided in Table A8. For selected outcomes the estimates are in Figure 10, all other estimates being in Appendix Figures A6 (for child outcomes) and A7 (for parent-level outcomes).

We see that twin birth rates decline significantly in each subgroup. Moreover, the magnitude of the decline is similar across all subgroups subject to the mandate, a pattern consistent with it resulting from the mandate rather than from selection or an omitted factor. This pervasive pattern is mirrored in indicators such as child birth weight and gestational age that directly reflect twinning.

The impact of SET on the mother's health, fertility and earnings shows some variation across subgroups. For instance, the result that IVF-users are more likely to continue fertility after the SET mandate is driven by women at first parity, and by women with higher education. A number of other interesting patterns are detailed in Appendix D. In general, we see significant declines in the high-frequency subgroups, which confirm that most IVF-users experienced the noted improvements.[39] Overall, the impacts of SET are fairly pervasive by markers of the demographics, health and education of women.

[39]Most outcomes show similar magnitudes across education and BMI categories, but some outcomes show no significant impact for women at parity 3 or higher, or among women under the age of 25, both of which are low-frequency categories in the treated (IVF) sample.

7.3 A back-of-the envelope cost-benefit analysis of the SET mandate

Given the benefits documented over a range of outcomes, and costs implicit in conducting SET rather than DET procedures, we seek to obtain a broad estimate of the reform's implications by conducting a back-of-the-envelope cost-benefit analysis. Full details of our calculations are provided in Appendix F, which is summarised here.

The financial costs of SET are equal to the costs of DET, as shown in the reimbursement register maintained by the Swedish Association of Local Authorities and Regions. However, we adjust the costs of SET (vs DET) upward to account for a 15 pp difference in the probability of a live birth between DET and SET with the first procedure, which is 42% vs 27% (McLernon et al., 2010). Recall that SET and DET have similar success rates cumulatively but not at the first attempt. We estimate short run benefits to include the costs of hospital nights and emergency C-sections averted, and increased post-birth earnings. We estimate long run benefits by using estimates available in the literature that allow us to project the reduced probability of low birth weight onto future lifetime earnings for affected children.

Taken together over the lifetime of the mother and child the benefits from mandating SET overwhelm its costs by a factor of 116. If we considered only short run benefits (that include mother's earnings in the 9 years following birth but exclude the earnings of children once they grow up) this ratio is smaller but still exceeds costs by a factor of 62. We under-estimate the short run benefits by virtue of ignoring the medical care costs averted on account of reduced morbidities among women and children that are not captured by hospital nights. We under-estimate the long run benefits to the extent that we (a) project forward only the economic gains to birth weight improvements and not, for instance, improvements in birth length or head circumference and (b) we summarize economic gains in earnings, rather than also accounting for the impacts of early life health improvements on other dimensions including cognitive attainment, employment and life expectancy which may have impacts beyond earnings on next generation productivity or on health care costs.[40]

8 Conclusion

Linking administrative data from several sources at the individual level to create longitudinal data for all births in Sweden during 1998-2007, we provide a comprehensive examination of causal effects of a 2003 reform that mandated single embryo transfer (SET) in IVF treatment, displacing the default of double embryo transfers (DET). We find that, after SET, women using IVF experience a sharply lower probability of twin birth, better child and maternal health and higher earnings in

[40]Using Swedish data, Bhalotra et al. (2017) demonstrate large impacts of an early child health interventon on later life chronic disease and life expectancy. Bhalotra et al. (2021) demonstrate impacts on cognitive attainment and employment in addition to earnings, also for Sweden and estimates for other countries are surveyed in Almond et al. (2018). A number of studies show that parental human capital has causal impacts on next generation human capital.

the nine years following birth. We estimate that the benefits flowing from SET vastly outweigh its costs. Our findings are important as IVF is now a key feature of the reproductive landscape and likely to continue to increase, especially as it becomes more readily accessible to women in poorer countries. Our results are more broadly relevant to the rising share of twin births among women not using IVF, driven by both delayed parenthood and improvements in maternal health.

References

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.

Aizer, A. (2010): "The Gender Wage Gap and Domestic Violence," *American Economic Review*, 100, 1847–59.

Aldén, L., A. Bjorklund, and M. Hammarstedt (2017): "Early Health and School Outcomes for Children with Lesbian Parents: Evidence from Sweden," IZA Discussion Papers 10616, Institute of Labor Economics (IZA).

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 (2018): "Childhood Circumstances and Adult Outcomes: Act II," *Journal of Economic Literature*, 56, 1360–1446.

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

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

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

Beemsterboer, S., R. Homburg, N. Gorter, R. Schats, P. Hompes, and C. Lambalk (2006): "The paradox of declining fertility but increasing twinning rates with advancing maternal age," *Human Reproduction*, 21, 1531–1532.

Bergh, C., A. T. Kjellberg, and P.-O. Karlström (2005): "Ett ägg vid provrörsbefruktning," *Läkartidningen*, 102, 3444–3449.

Berniell, I., L. Berniell, D. de la Mata, M. Edo, and M. Marchionni (2021): "Gender gaps in labor informality: The motherhood effect," *Journal of Development Economics*, 150, 102599.

Bhalotra, S. and D. Clarke (2019): “Twin Birth and Maternal Condition,” *Review of Economics and Statistics*, 101, 853–864.

——— (2020): “The twin instrument: Fertility and human capital investment,” *Journal of the European Economic Association*, 18, 3090–3139.

Bhalotra, S., D. Clarke, H. Mühlrad, and M. Palme (2019): “Multiple Births, Birth Quality and Maternal Labor Supply: Analysis of IVF Reform in Sweden,” Working Paper Series 1289, Research Institute of Industrial Economics.

Bhalotra, S., D. Hollywood, and A. Venkataramani (2022): “Fertility, health endowments and returns to human capital: Quasi experimental evidence from 20th century america,” Tech. rep.

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., M. Karlsson, T. Nilsson, and N. Schwarz (2021): “Infant Health, Cognitive Performance and Earnings: Evidence from Inception of the Welfare State in Sweden,” *The Review of Economics and Statistics*, 1–46.

Bhalotra, S. R. and D. Clarke (2022): “Analysis of Twins,” IZA Discussion Papers 15609, Institute of Labor Economics (IZA).

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

Bitler, M. P. (2008): “Effects of Increased Access to Infertility Treatment on Infant Health Outcomes: Evidence from Twin Births,” mimeo.

Bitler, M. P. and L. Schmidt (2012): “Utilization of Infertility Treatments: The Effects of Insurance Mandates,” *Demography*, 49, 125–149.

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

Conley, D. and R. Glauber (2005): “Parental Educational Investment and Children’s Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variations in Fertility,” Working Paper 11302, National Bureau of Economic Research.

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

Cáceres-Delpiano, J. (2006): “The Impacts of Family Size on Investment in Child Quality,” *Journal of Human Resources*, XLI, 738–754.

de Chaisemartin, C. and X. D’Haultfœuille (2020): “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects,” *American Economic Review*, 110, 2964–96.

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.

De Sutter, P., J. Van der Elst, T. Coetsier, and M. Dhont (2003): "Single embryo transfer and multiple pregnancy rate reduction in IVF/ICSI: a 5-year appraisal," *Reproductive BioMedicine Online*, 6, 464–469.

Ekechi, C. (2021): "Addressing inequality in fertility treatment," *The Lancet*, 398, 645–646.

Eskew, A. M. and E. S. Jungheim (2017): "A history of developments to improve in vitro fertilization," *Missouri medicine*, 114, 156.

European Society of Human Reproduction and Embryology (2018): "More than 8 million babies born from IVF since the world's first in 1978: European IVF pregnancy rates now steady at around 36 percent, according to ESHRE monitoring," Tech. rep., ScienceDaily. ScienceDaily, 3 July 2018. www.sciencedaily.com/releases/2018/07/180703084127.htm. Accessed December, 2021.

Gardner, D. K., E. Surrey, D. Minjarez, A. Leitz, J. Stevens, and W. B. Schoolcraft (2004): "Single blastocyst transfer: a prospective randomized trial," *Fertility and sterility*, 81, 551–555.

Gerris, J., D. De Neubourg, K. Mangelschots, E. Van Royen, M. Van de Meerssche, and M. Valkenburg (1999): "Prevention of twin pregnancy after in-vitro fertilization or intracytoplasmic sperm injection based on strict embryo criteria: a prospective randomized clinical trial," *Human Reproduction*, 14, 2581–2587.

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.

Gershoni, N. and C. Low (2021a): "Older Yet Fairer: How Extended Reproductive Time Horizons Reshaped Marriage Patterns in Israel," *American Economic Journal: Applied Economics*, 13, 198–234.

——— (2021b): "The power of time: The impact of free IVF on Women's human capital investments," *European Economic Review*, 133, 103645.

Geyter, C. D. (2021): "Single embryo transfer in all infertile couples treated with assisted reproduction produces excellent results and avoids multiple births," *Swiss Medical Weekly*.

Goodman-Bacon, A. (2021): "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*.

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

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

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

Karimi, A. (2014): "The spacing of births and women's subsequent earnings: Evidence from a natural experiment," Working Paper 2014:18, Uppsala.

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.

Kjellberg, A. T., P. Carlsson, and C. Bergh (2006): "Randomized single versus double embryo transfer: obstetric and paediatric outcome and a cost-effectiveness analysis," *Human Reproduction*, 21, 210–216.

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

Kleven, H. J. and C. T. Kreiner (2003): "The Marginal Cost of Public Funds in OECD Countries. Hours of Work Versus Labor Force Participation," CESifo Working Paper Series 935, CESifo.

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

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

Levine, H., N. Jørgensen, A. Martino-Andrade, J. Mendiola, D. Weksler-Derri, I. Mindlis, R. Pinotti, and S. H. Swan (2017): "Temporal trends in sperm count: a systematic review and meta-regression analysis," *Human Reproduction Update*, 23, 646–659.

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.

Lundberg, S. and E. Rose (2002): "The Effects Of Sons And Daughters On Men's Labor Supply And Wages," *The Review of Economics and Statistics*, 84, 251–268.

Lundberg, S. J., R. A. Pollak, and T. J. Wales (1997): "Do Husbands and Wives Pool Their Resources? Evidence from the United Kingdom Child Benefit," *The Journal of Human Resources*, 32, 463–480.

Lundborg, P., E. Plug, and A. W. Rasmussen (2017): "Can Women Have Children and a Career? IV Evidence from IVF Treatments," *American Economic Review*, 107, 1611–37.

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.

Martikainen, H., A. Tiitinen, C. Tomás, J. Tapanainen, M. Orava, L. Tuomivaara, S. Vilska, C. Hydén-Granskog, and O. Hovatta (2001): "One versus two embryo transfer after IVF and ICSI: a randomized study," *Human Reproduction*, 16, 1900–1903.

McElroy, M. B. and M. J. Horney (1981): "Nash-Bargained Household Decisions: Toward a Generalization of the Theory of Demand," *International Economic Review*, 22, 333–349.

McLernon, D. J., K. Harrild, C. Bergh, M. J. Davies, D. De Neubourg, J. Dumoulin, J. Gerris, J. A. Kremer, H. Martikainen, B. Mol, et al. (2010): "Clinical effectiveness of elective single versus double embryo transfer: meta-analysis of individual patient data from randomised trials," *Bmj*, 341.

Monden, C., G. Pison, and J. Smits (2021): "Twin Peaks: more twinning in humans than ever before," *Human Reproduction*, 36, 1666–1673.

National Board of Health and Welfare (2020): "Fertility treatments in Sweden," National report.

OECD (2021): "OECD Family Database," Indicator SF2.3 Online resource: . Accessed December 2021.

Oster, E. (2019): "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 37, 187–204.

Pew Research Center (2021): "A third of U.S. adults say they have used fertility treatments or know someone who has," Online resource: https://www.pewresearch.org/fact-tank/2018/07/17/a-third-of-u-s-adults-say-they-have-used-fertility-treatments-or-know-someone-who-has/. Accessed December 2021.

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.

Pison, G. and A. V. D'Addato (2006): "Frequency of Twin Births in Developed Countries," *Twin Research and Human Genetics*, 9, 250–259.

Qian, N. (2017): "The Effect of China's One Child Policy on Sex Selection, Family Size and the School Enrollment of Daughters," *Gender and Development*, Forthcoming.

Rambachan, A. and J. Roth (2020): "An Honest Approach to Parallel Trends," Working paper, Harvard University, Department of Economics.

Romano, J. P. and M. Wolf (2005): “Stepwise Multiple Testing as Formalized Data Snooping,” *Econometrica*, 73, 1237–1282.

Rosenzweig, M. R. and J. Zhang (2009): “Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight and China’s ”One-Child” Policy,” *The Review of Economic Studies*, 76, 1149–1174.

Rossin-Slater, M. (2013): “WIC in your neighborhood: New evidence on the impacts of geographic access to clinics,” *Journal of Public Economics*, 102, 51–69.

Roth, J. (2021): “Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends,” Available at: https://jonathandroth.github.io/assets/files/roth_pretrends_testing.pdf, Conditionally accepted, American Economic Review: Insights.

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.

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*.

Sun, L. and S. Abraham (2020): “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects,” *Journal of Econometrics*.

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.

Tiitinen, A., L. Unkila-Kallio, M. Halttunen, and C. Hyden-Granskog (2003): “Impact of elective single embryo transfer on the twin pregnancy rate,” *Human Reproduction*, 18, 1449–1453.

van Montfoort, A. P., J. C. Dumoulin, J. A. Land, E. Coonen, J. G. Derhaag, and J. L. Evers (2005): “Elective single embryo transfer (eSET) policy in the first three IVF/ICSI treatment cycles,” *Human Reproduction*, 20, 433–436.

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.

Wennerholm, U.-B., S. Saltvedt, A. Wessberg, M. Alkmark, C. Bergh, S. B. Wendel, H. Fadl, M. Jonsson, L. Ladfors, V. Sengpiel, J. Wesström, G. Wennergren, A.-K. Wikström, H. Elden, O. Stephansson, and H. Hagberg (2019): "Induction of labour at 41 weeks versus expectant management and induction of labour at 42 weeks (SWEdish Post-term Induction Study, SWEPIS): multicentre, open label, randomised, superiority trial," *British Medical Journal*, 367.

Wikland, M. (2005): "Provrörsbefruktning gav avstamp för en nästan ofattbar utveckling," *Läkartidningen*.

Figures and Tables

Table 1: The Impact of SET on Fertility

	Twin Birth	# Births within 9 years	No future births
SET reform	-0.123***	0.074***	-0.072***
	[0.005]	[0.010]	[0.007]
IVF	0.175***	-0.106***	0.101***
	[0.005]	[0.008]	[0.006]
Scaled Impact of SET	-0.703	-0.694	-0.707
Observations	895,336	895,336	895,336
Mean of Dep. Var. IVF	0.190	0.579	0.536
Mean of Dep. Var. Full	0.016	0.686	0.472
Uncorrected p-value (SET)	0.000	0.000	0.000
Corrected p-value (SET)	< 0.001	< 0.001	< 0.001

Notes to Table 1: Each column presents a separate two-way fixed effect regression estimating the impact of the SET reform on fertility outcomes. Estimated date of conception fixed effects are included in all regressions. The scaled impact of SET refers to the proportional impact of the SET reform compared to the difference between IVF and non-IVF mothers. This is the ratio of the coefficient in row 1 to the coefficient in row 2. Corrected p-values are calculated using Holm's Family Wise Error Rate correction across all models estimated in Tables 1-4. Standard errors are clustered by mother. * p<0.1, ** p<0.05, *** p<0.01.

Table 2: The Impact of SET on Child Health

Panel A	Birth weight	Gestational age (weeks)	Hospital nights (age 0)	Hospital nights (aged 1-8)
SET reform	194.673***	0.611***	-1.787***	-0.126
	[12.043]	[0.048]	[0.251]	[0.123]
IVF	-319.311***	-1.090***	2.868***	0.334***
	[9.782]	[0.040]	[0.213]	[0.097]
Scaled Impact of SET	-0.610	-0.560	-0.623	-0.378
Observations	905,473	908,232	903,605	908,232
Mean of Dep. Var. IVF	3182.140	38.269	4.724	1.335
Mean of Dep. Var. Full	3539.100	39.315	1.708	1.064
Uncorrected p-value (SET)	0.000	0.000	0.000	0.303
Corrected p-value (SET)	< 0.001	< 0.001	< 0.001	1.00

Panel B	APGAR < 7	Severe neo-natal morbidity	Infant mortality	Under 5 mortality
SET reform	-0.004**	-0.031***	-0.001	-0.001
	[0.002]	[0.005]	[0.001]	[0.001]
IVF	0.006***	0.055***	0.003***	0.001**
	[0.002]	[0.004]	[0.001]	[0.001]
Scaled Impact of SET	-0.751	-0.556	-0.416	-0.760
Observations	900,261	908,232	908,232	908,232
Mean of Dep. Var. IVF	0.021	0.148	0.009	0.002
Mean of Dep. Var. Full	0.012	0.075	0.006	0.001
Uncorrected p-value (SET)	0.028	0.000	0.354	0.118
Corrected p-value (SET)	0.280	< 0.001	1.00	0.826

Notes to Table 2: Each column presents a separate two-way FE regression estimating the impact of the SET reform on neonatal and child health outcomes. All data are obtained from the Swedish Medical Birth Registry and the Swedish National Patient Registry covering all births for the time period 1998-2007. In the case of longer-term health outcomes (under 5 mortality and all hospitalization between ages 1-8), these are observed in the Swedish National Patient Registry following children up until (a maximum of) 2016. Estimated date of conception fixed effects are included in all regressions. The scaled impact of SET refers to the proportional impact of the SET reform compared to the difference between IVF and non-IVF mothers. This is the ratio of the coefficient in row 1 to the coefficient in row 2. Corrected p-values are calculated using Holm's Family Wise Error Rate correction across all models estimated in Tables 1-4. Standard errors are clustered by mother. * p<0.1, ** p<0.05, *** p<0.01.

Table 3: The Impact of SET on Maternal Health

	Maternal morbidity	Hospital nights childbirth	Emergency C-section
SET reform	-0.011*	-0.633***	-0.015**
	[0.006]	[0.113]	[0.006]
IVF	0.055***	0.997***	0.042***
	[0.005]	[0.099]	[0.004]
Scaled Impact of SET	-0.200	-0.635	-0.346
Observations	895,336	874,814	895,336
Mean of Dep. Var. IVF	0.228	5.214	0.171
Mean of Dep. Var. Full	0.132	3.603	0.086
Uncorrected p-value (SET)	0.086	0.000	0.011
Corrected p-value (SET)	0.688	< 0.001	0.149

Notes to Table 3: Each column presents a separate two-way fixed effect regression estimating the impact of the SET reform on maternal health or hospitalization measures. Estimated date of conception fixed effects are included in all regressions. The scaled impact of SET refers to the proportional impact of the SET reform compared to the difference between IVF and non-IVF mothers. This is the ratio of the coefficient in row 1 to the coefficient in row 2. Corrected p-values are calculated using Holm's Family Wise Error Rate correction across all models estimated in Tables 1-A11. Standard errors are clustered by mother. * p<0.1, ** p<0.05, *** p<0.01.

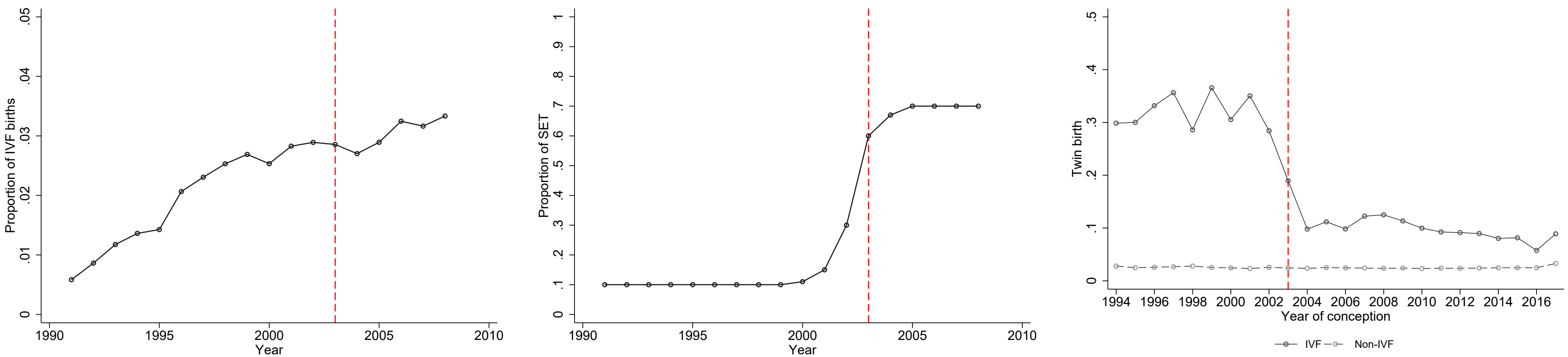
Table 4: The Impact of SET on Labour Market Outcomes

	Mothers		Fathers	
	Wage Earnings	Sickness Benefits	Wage Earnings	Sickness Benefits
SET reform	10872***	-665***	-2192	-166
	(2061)	(207)	(5046)	(205)
IVF	-1409	1656***	22728***	438**
	(1526)	(173)	(4151)	(178)
Scaled Impact of SET (vs. baseline)	0.056	-0.082	-0.005	-0.037
Observations	893,747	893,747	893,793	893,793
Mean of Dep. Var. IVF	192983	8127	400782	4481
Mean of Dep. Var. Full	162822	6845	338141	4413
Uncorrected p-value (SET)	0.000	0.001	0.664	0.418
Corrected p-value (SET)	< 0.001	0.017	1.00	1.00

Notes to Table 4: Each column presents a separate two-way FE estimate of the impact of the SET reform on maternal or paternal labour market outcomes. Wage earning and sickness benefits refer to averages over the 9 years following each mother or father's birth, and are generated by following parents up to (a maximum of) 2016 in the Longitudinal integration database for health insurance and labour market studies (LISA) register. Estimated date of conception fixed effects are included in all regressions. The scaled impact of SET (vs. baseline) refers to the proportional impact of the estimate of the SET reform (row 1), compared to the mean of the dependent variable among IVF users at baseline. Corrected p-values are calculated using Holm's Family Wise Error Rate correction across all models estimated in Tables 1-4. Standard errors are clustered by mother. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Figure 1: Trends in SET and proportion of IVF births

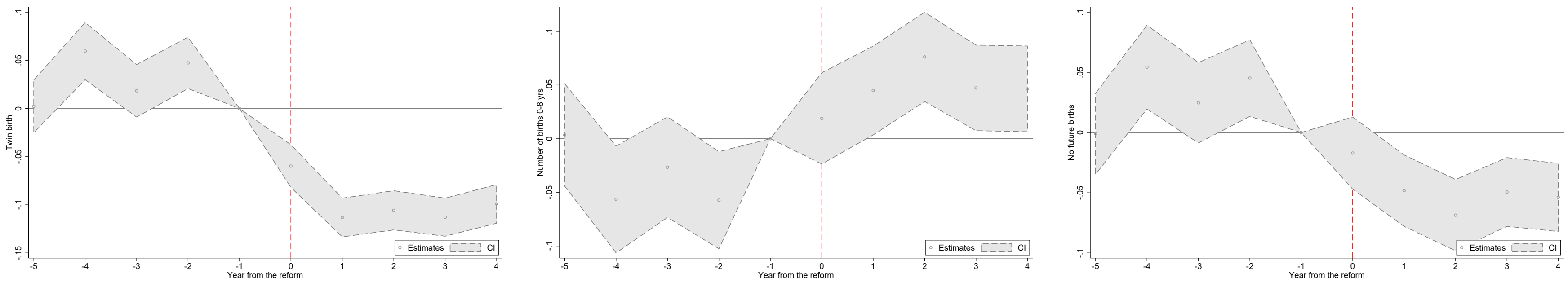
(a) Proportion of IVF births | **(b)** Single embryo transfer | **(c)** Trends in twin rates



Notes: Annual trends in the proportion of IVF and the proportion of SET births are based on aggregate data collected from annual reports by the Swedish National Board of Health and Welfare. Rates of twin births are based on micro-level data from the Swedish Medical Birth Registry recording all pregnancies over at least 22 weeks of gestation. Proportions in panel (a) refer to the proportion among all births. Proportions in panel (b) refer to the proportion among all IVF birth and proportions in panel (c) refer to the proportion among IVF (solid line) or non-IVF (dashed line) births. The red vertical line indicates the year of the SET reform.

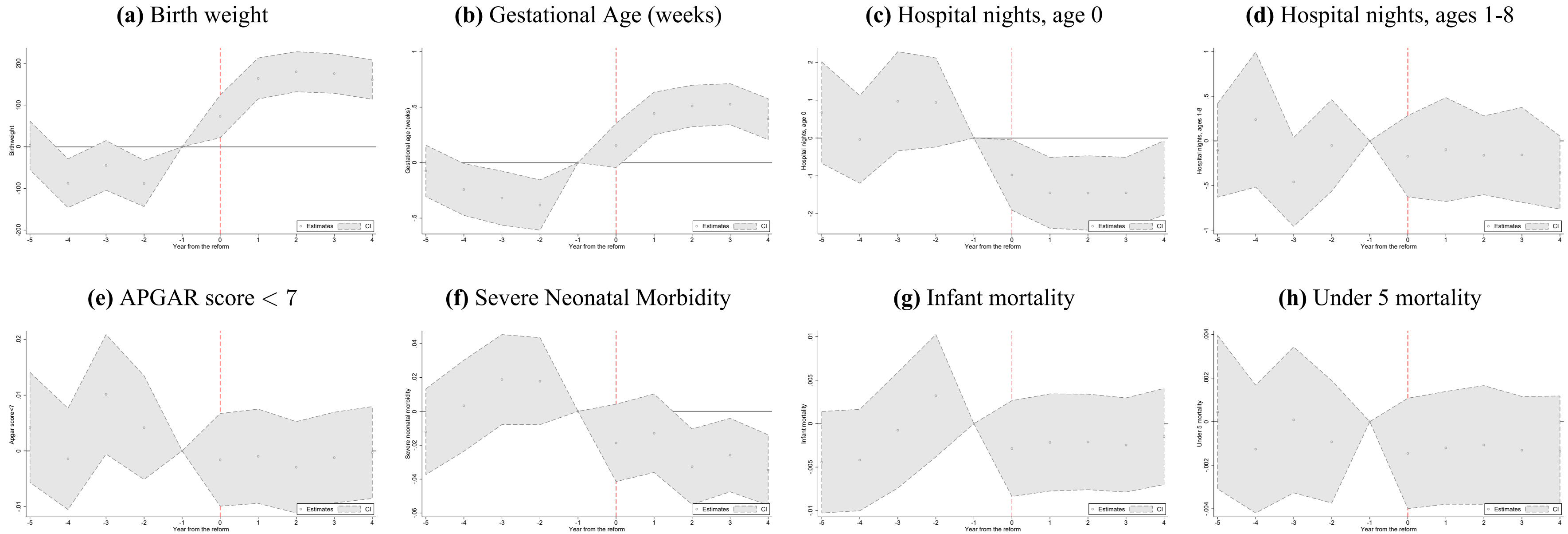
Figure 2: Event studies: The Impact of SET on Fertility

(a) Twin birth **(b)** Number of future births within 9 years **(c)** No more births



Notes: Event studies are analogous to single coefficient regressions presented in Table 1, however here estimate full leads and lags to the exposure to the SET reform interacted with a mother's IVF status, as per specification 1. Leads and lags refer to cohorts of mothers giving birth up to 5 years pre and 4 years post-SET reform. The red-vertical line represents the year of the SET reform, and year -1 is the omitted reference period. Standard errors are clustered by mother.

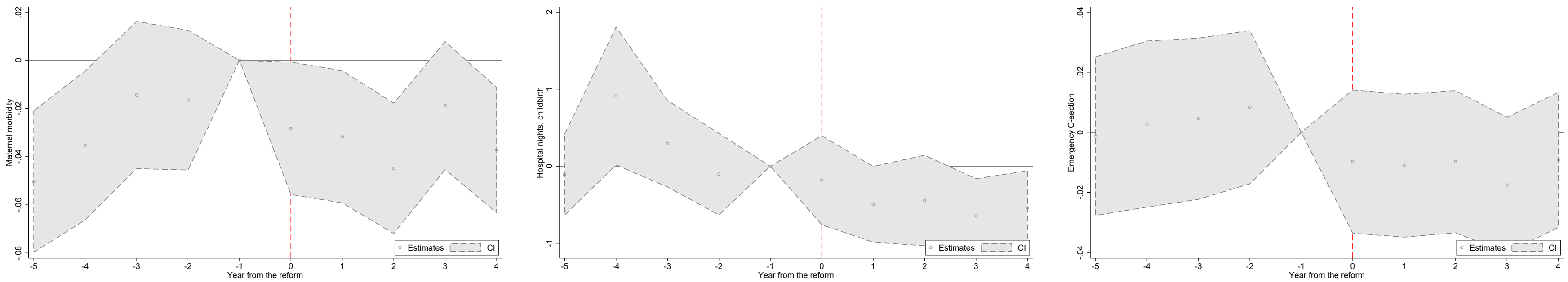
Figure 3: Event studies: The Impact of SET on Neonatal and Child Health



Notes: Event studies are analogous to single coefficient regressions presented in Table 2, however here estimate full leads and lags to the exposure to the SET reform interacted with each child's mother's IVF status, as per specification 1. Leads and lags refer to cohorts of children born up to 5 years pre and 4 years post-SET reform. The red-vertical line represents the year of the SET reform, and year -1 is the omitted reference period. Standard errors are clustered by mother.

Figure 4: Event studies: The Impact of SET on Maternal Health

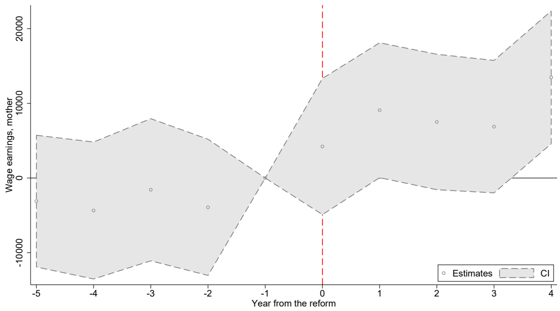
(a) Maternal Morbidity Index **(b)** Hospital nights **(c)** Emergency-C-section



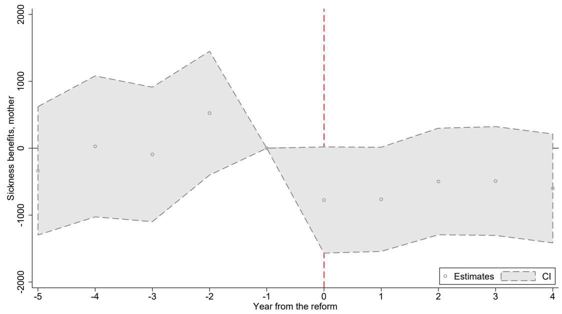
Notes: Event studies are analogous to single coefficient regressions presented in Table 3, however here estimate full leads and lags to the exposure to the SET reform interacted with a mother's IVF status, as per specification 1. Leads and lags refer to cohorts of mothers giving birth up to 5 years pre and 4 years post-SET reform. The red-vertical line represents the year of the SET reform, and year -1 is the omitted reference period. Standard errors are clustered by mother.

Figure 5: Event studies: The Impact of SET on Labor Market Outcomes

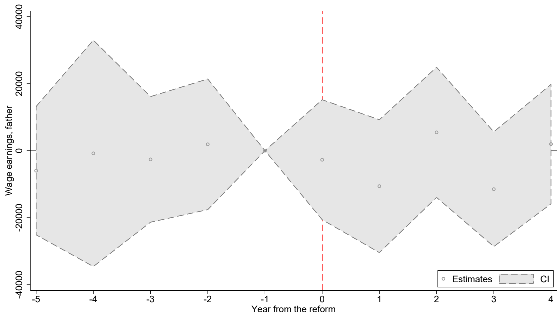
(a) Wage earnings, mother



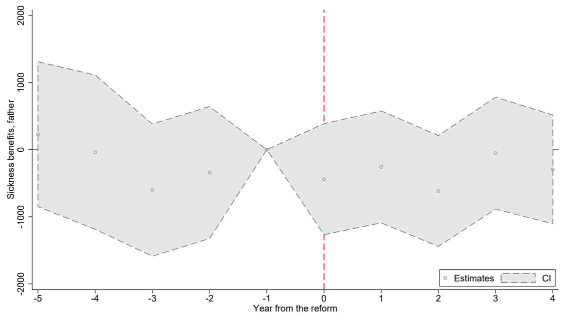
(b) Income from sickness leave, mother



(c) Wage earnings, father

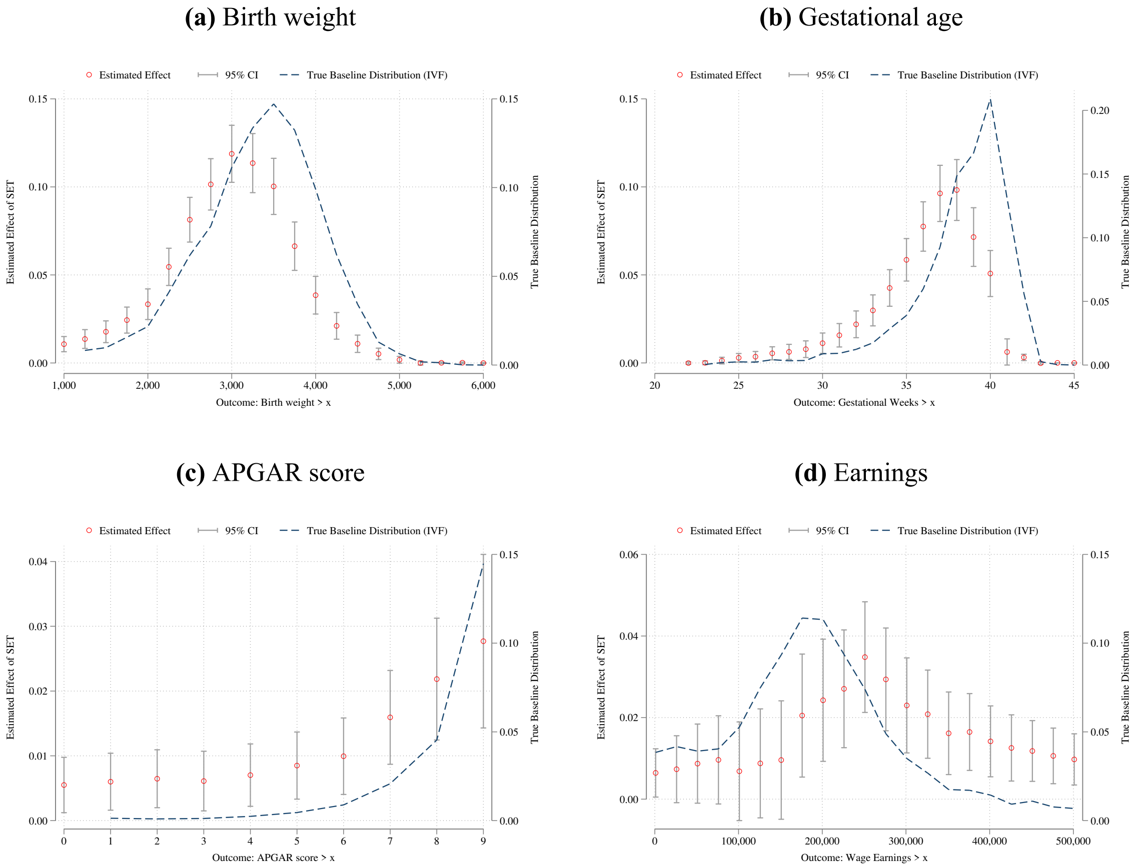


(d) Income from sickness leave, father



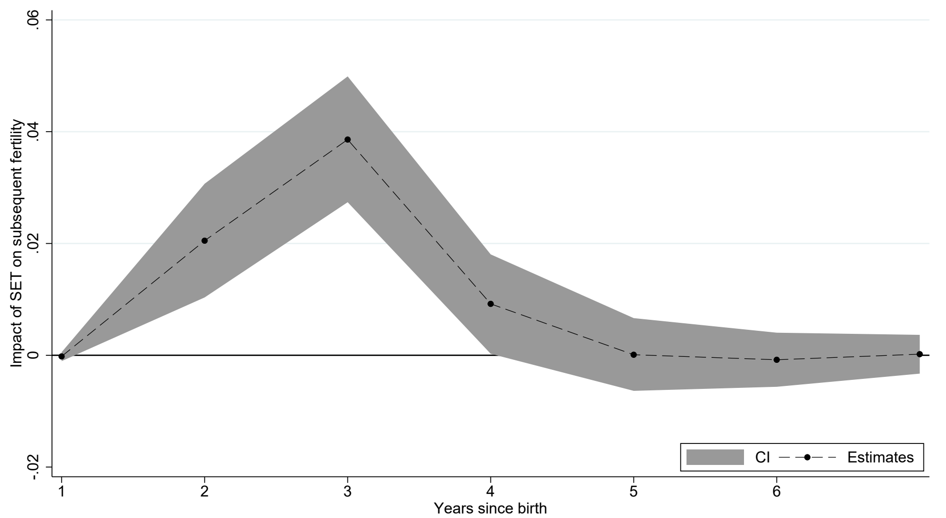
Notes: Event studies are analogous to single coefficient regressions presented in Table 4, however here estimate full leads and lags to the exposure to the SET reform interacted with each mother's IVF status, as per specification 1. Leads and lags refer to cohorts of mothers or fathers of children born up to 5 years pre- and 4 years post-SET reform. The red-vertical line represents the year of the SET reform, and year -1 is the omitted reference period. Standard errors are clustered by mother or father.

Figure 6: Distributional Impacts of the SET Reform



Notes: Distributional impacts of the SET reform are considered over child outcomes (panels (a)-(c)) and mother's labour market outcomes (panel (d)). In each case, blue dashed lines present baseline distributions of actual outcomes in the population of IVF users. Point estimates (red circles) and 95% CIs (grey error bars) refers to the estimated impact of SET from a difference-in-differences regression where the outcome is whether the birth exceeds the particular threshold indicated on the horizontal axis (panels (a)-(c)) or whether the mother's average earnings exceed the particular average earnings threshold indicated on the horizontal axis of panel (d). Specifically, each point and CI refers to $\hat{\beta}_1^k$ from the following specification:
$Pr(Y_{it} > k) = \alpha^k + \beta_1^k(PostSET \times IVF)_{it} + \beta_2^k IVF_i + \boldsymbol{X}_{it}^k \delta + \alpha_c^k + \pi_t^k + \varepsilon_{it}^k$
estimated using a linear probability model for each cut-point k documented on each plot, and is thus interpreted as the marginal change in the likelihood of exceeding particular distributional points of interest among IVF mother's post SET, compared with IVF mother's pre-SET in a double-difference framework with non-IVF mothers. All other details follow those in regressions estimated in Tables 2-4. Standard errors are clustered by mother.

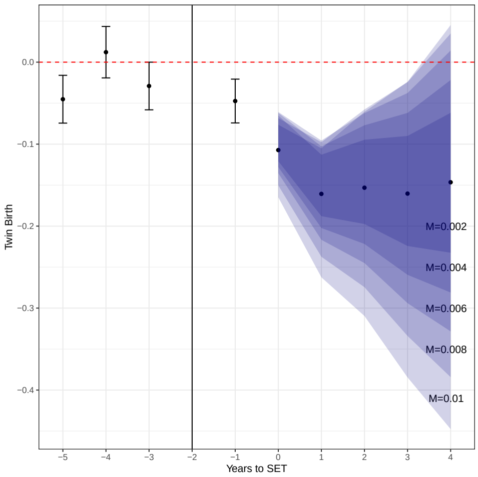
Figure 7: Impact of SET on future fertility and birth spacing



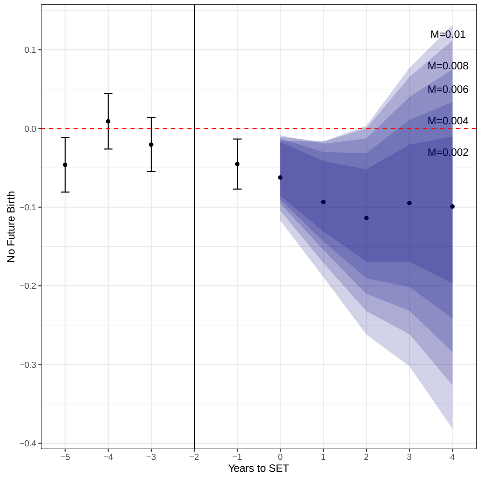
Notes: Each coefficient and standard error are obtained from a difference-in-differences regression of having been exposed to the SET reform on the probability of having an additional birth in the years following all births observed in our data (the "index births"). Each index birth is obtained from the Swedish Medical Birth Registry for the time period 1998-2007, with following births occurring up to 9 years post-index births, (therefore followed up until a maximum of 2016). Standard errors are clustered by mother, and 95% confidence intervals (CIs) are plotted.

Figure 8: Honest DiD Bounds – Partial Identification Relaxing Parallel Trend Assumptions

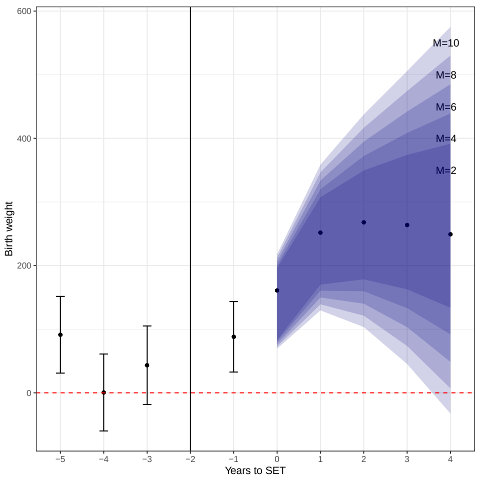
(a) Twin birth



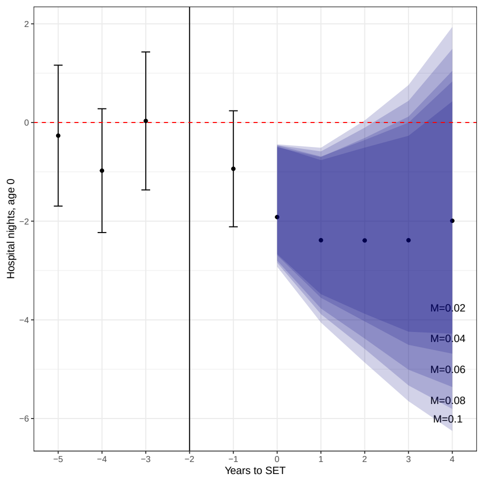
(b) No future births



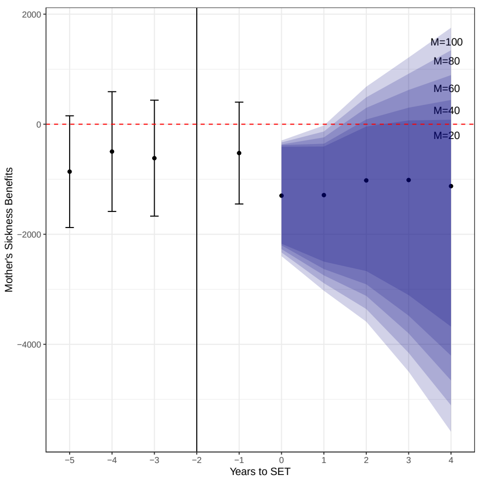
(c) Birth weight



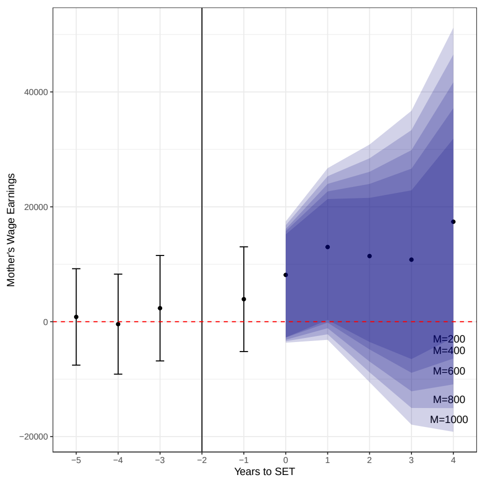
(d) Hospital nights infancy



(e) Sickness benefits mother



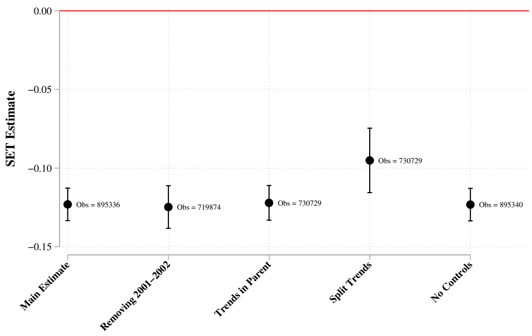
(f) Wage earnings mother



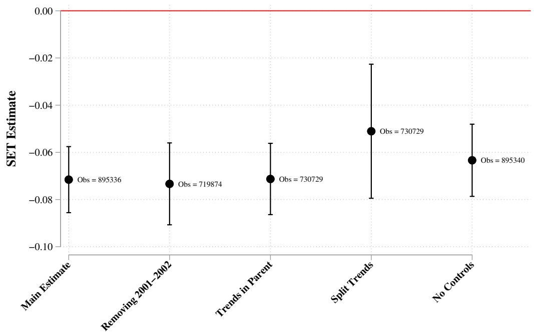
Notes: "Honest Difference-in-Differences" models are estimated following Rambachan and Roth (2020). Here rather than assuming parallel trends between IVF and non-IVF mothers in the post-SET period, the counterfactual is assumed to follow trends between IVF and non-IVF mothers observed in the pre-SET period. These trends are additionally allowed to diverge by up to M units in each period, where M is indicated in each shaded confidence interval plotted. In each case, solid shaded areas present 95% confidence bounds. Black points represent original event study estimates, and black error bars prior to period 0 refer to original (event study) estimates used to model pre-trends. Given some roll-out in year -1, pre-trends are estimated from years -5 to -2, with year -2 being the omitted baseline year.

Figure 9: Estimates Based on Alternative Models and Samples

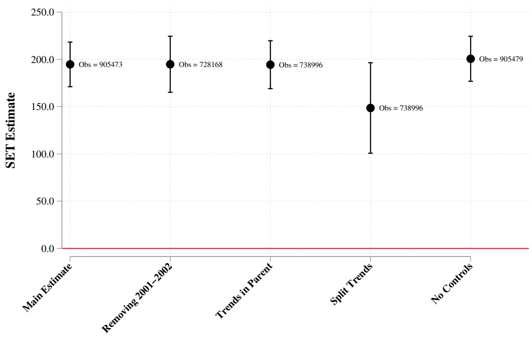
(a) Twin birth



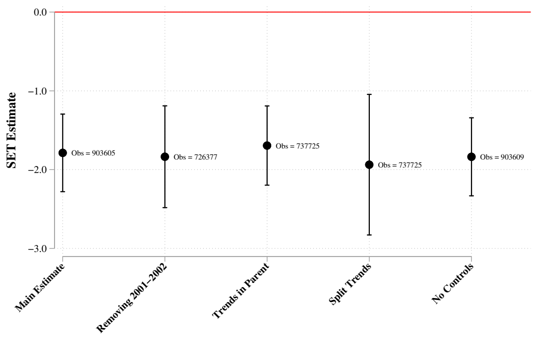
(b) No future births



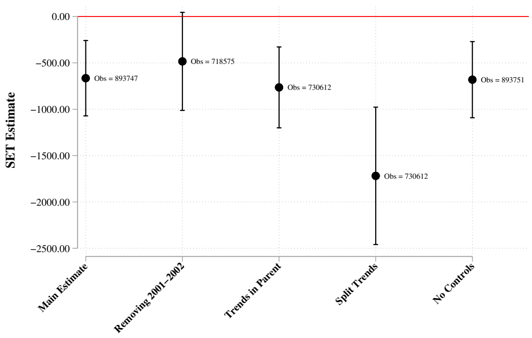
(c) Birth weight



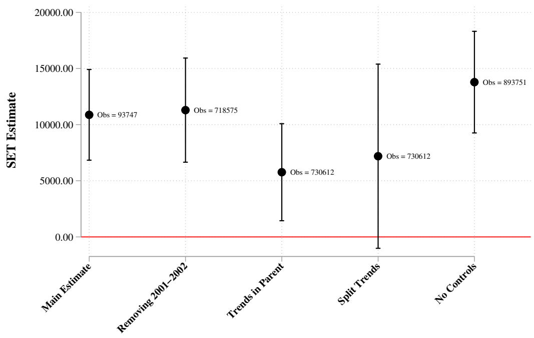
(d) Hospital nights infancy



(e) Sickness benefits mother

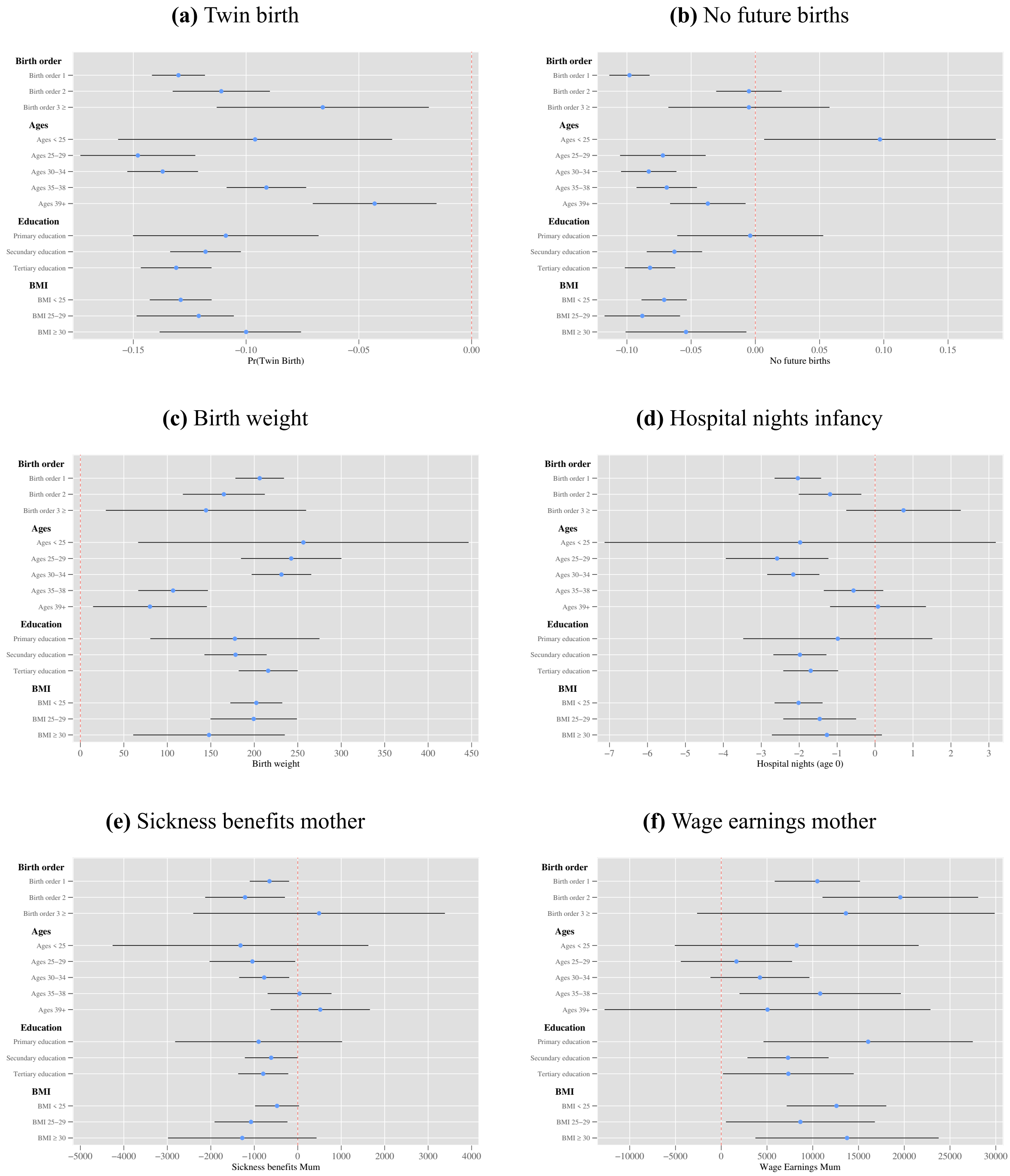


(f) Wage earnings mother



Notes: Each plot estimates alternative specifications for outcomes examined in Table 1 of the paper. "Main Estimate" reproduces the estimate from Table 1 for comparability. "Removing 2001-2002" implements a Donut DD model where the pre-treatment years of 2001 and 2002 are removed. "Trends in Parent" refer to models including trends in all parental characteristics indicated in the Data section of the paper, while "Split Trends" refers to models including these trends separately for IVF and non-IVF parents. "No Controls" is a baseline model without any controls. Black circles present point estimates and error bars present 95% confidence intervals. All estimates and CIs are generated from DD models, with standard errors clustered by mother.

Figure 10: Reform Heterogeneity



Notes: Each plot documents heterogeneity in reform impacts by birth order, maternal age, education and BMI. Each point estimate (solid circle) and 95% confidence interval (error bars) plots the estimate of the impact of the SET reform estimated from DD models where the estimation sample consists only of individuals meeting the criteria indicated in the vertical axes. All other details follow those from equation 2 of the text.

Table 5: Bounds Estimates Based on Unobservable Selection in SET

Panel A: Child Health Outcomes

	Birth weight	Gestational age (weeks)	Hosp. nights (age 0)	Hosp. nights (aged 1-8)	APGAR < 7	Severe neonate morbidity	Infant mortality	Under 5 mortality
Baseline Effect	200.592***	0.619***	-1.837***	-0.135	-0.005**	-0.032***	-0.001	-0.001
Standard Error	(12.161)	(0.048)	(0.253)	(0.123)	(0.002)	(0.005)	(0.001)	(0.001)
R^2	0.005	0.004	0.002	0.000	0.000	0.002	0.000	0.000
Controlled Effect	194.673***	0.611***	-1.787***	-0.126	-0.004**	-0.031***	-0.001	-0.001
Standard Error	(12.043)	(0.048)	(0.251)	(0.123)	(0.002)	(0.005)	(0.001)	(0.001)
R^2	0.071	0.012	0.008	0.001	0.003	0.014	0.002	0.001
Bounds and Maximal Selection								
$\tilde{\delta}$ for $\beta = 0$	97.23	137.63	86.28	37.78	79.84	61.13	113.98	156.96
Given R_{max}								
Identified set	[192.755, 194.673]	[0.607, 0.611]	[-1.787, -1.769]	[-0.126, -0.123]	[-0.004, -0.004]	[-0.031, -0.030]	[-0.001, -0.001]	[-0.001, -0.001]
Extreme Bound with $R_{max} = 1$	$111.36^{\dagger}$	-0.377	6.479	14.86	0.379	$-0.007^{\dagger}$	—	—

Panel B: Mother's Outcomes (Fertility, Health, Labour Market)

	Twin Birth	# Births in 9 years	No future births	Maternal morbidity	Hospital nights birth	Emergency C-section	Wage Earnings	Sickness Benefits
Baseline Effect	-0.123***	0.063***	-0.063***	-0.013*	-0.673***	-0.017***	1.4e+04***	-681.324***
Standard Error	(0.005)	(0.011)	(0.008)	(0.006)	(0.113)	(0.006)	(2309.013)	(209.502)
R^2	0.019	0.001	0.001	0.002	0.006	0.002	0.022	0.002
Controlled Effect	-0.123***	0.074***	-0.072***	-0.011*	-0.633***	-0.015**	1.1e+04***	-665.218***
Standard Error	(0.005)	(0.010)	(0.007)	(0.006)	(0.113)	(0.006)	(2060.700)	(207.307)
R^2	0.020	0.258	0.307	0.023	0.028	0.030	0.309	0.033
Bounds and Maximal Selection								
$\tilde{\delta}$ for $\beta = 0$	52.31	-22.58	-29.04	20.58	42.60	18.38	11.53	126.43
Given R_{max}								
Identified set	[-0.123, -0.122]	[0.074, 0.077]	[-0.074, -0.072]	[-0.011, -0.010]	[-0.633, -0.619]	[-0.015, -0.014]	[9930.649, 1.1e+04]	[-665.218, -660.055]
Extreme Bound with $R_{max} = 1$	$-0.025^{\dagger}$	$0.105^{\dagger}$	$-0.092^{\dagger}$	0.082	0.862	0.054	$4799.30^{\dagger}$	$162.84^{\dagger}$

Notes: Bounds and maximal degree of unobservable selection are presented based on Oster (2019). In each panel, baseline effect reports estimates of the impact of SET based on DD models with state and time fixed effects, but no time-varying controls. Then controlled effect presents identical models, however no including observable controls. Estimate of δ for $\beta = 0$ describe how much more important unobservables would need to be than observables to drive the estimated effect of SET to 0, assuming that the maximum R^2 is 1.3 times the R^2 in controlled models. Identified set refers to bounds following Oster (2019) under assumptions that unobservables are equally as important as observables ($\delta = 1$), and that the maximum R^2 is 1.3 times the R^2 in controlled models. Finally, extreme bounds refers to the estimate on SET if $\delta = 1$ and the maximum R^2 is instead set to 1. In this case, the other side of the bounds is identical to the controlled effect. Extreme bounds are not presented for infant mortality, as the R^2 is so low to render such bounds useless. * p<0.1, ** p<0.05, *** p<0.01. † Extreme bounds are informative of the sign of the estimate.

Online Appendices for:

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

Sonia Bhalotra & Damian Clarke & Hanna Mühlrad & Mårten Palme

Table of Contents

A Appendix Figures and Tables

Table A1: Classification of Maternal Morbidity Events

Classification	ICD-10 and KVÅ Codes	MBR Variable Name
Maternal Morbidity		
Maternal deaths	O959 O960 O961 O969 O970 O971 O979	
Non-planned hysterectomy	O822, MCA33, MCA30, LCD00, LCD96+ZXD00	
Blood transfusion + blood loss >1000 ml	(DR029, DR033, DR036, DR038, V9209 +	
or haemorrhage with coagulopathy	O678, O720, O721A, O721B, O721X)	
	or O723 or O670	
Other surgical interventions such as uterine or vaginal tamponade, uterine compression sutures, embolization, re-operation due to complications	MBB10 DP015 MCW96 KCH00 MWA00 MWB00 MWC00 MWC01 MWD00 MWE00 MWE01	
	MWE02 MWF00 MWF01 MWW96 MWW97 MWW98	
Maternal venous thromboembolism	O871 O873 O878 O880 O881 O882 O883 O888	
Maternal sepsis	O859 R572 A410 A411 A412 A413 A414 A415 A418 A419	
Third degree perineal injury with over 50% of the external anal sphincter torn	O702 O702C 0702D 0702E 0702F 0702X + MBR-variable	SFINKTER
Fourth degree perineal injury (including the rectal mucosa)	0703 + MBR-variable	REKTUM
Injuries during delivery and surgery:	O711 O715	
Uterine rupture, bladder or urethra	S355 S357 S358 S359 S364 S365 S368 S371 S372	
Arteria iliaca interna, intestines, urinary bladder and urethra		
Anaesthesia complications	O890 O891 O892 O896 O903 O740 O741 O742 O743 O747 O751 O754	
Eclampsia	O15 O150 O151 O152 O159	
Post-partum depression	F32 F320 F321 F322 F323 F32A F323W F328 F329 F53 F530 F531 F538 F539	
Placenta complications	O430 O430A O430B O430W O430X O431 O432 O432A O432 O432X O438 O43	
Complications associated with multiple births	O310 O311 O312 O312A O312B O312X O318	
Wound rupture	O900 O901 O902	
Intensive care	ZV049	
Cervical lacerations	O713 MBC00	
Preeclampsia/gestational hypertension	O139 O14 O140 O141 O141A O141B O141X O142 O149	
Gestational Diabetes Mellitus	O244 O244A O244B	
Chorioamnionitis	O411 R572	
Wound infection	O860	
UTI	O862 O863	
Endometritis	O859	

Notes: Maternal morbidity events are documented in the Swedish Medical Birth Register (MFR) and defined using medical diagnoses classified according to the International Classification of Diseases 10 (ICD-10) and procedures (using KVÅ-codes) during pregnancy and delivery. The morbidity events listed in this table are defined by the ICD-10 and KVÅ codes presented in column 2 and/or by the MFR-variables presented in column 3.

Table A2: Classification of Severe Neonatal Morbidity Events

Classification	ICD-10 Code	MBR Variable Name
Severe Neonatal Morbidity		
Stillbirth (ante- and intrapartum)		DODFOD
Asphyxia at delivery, pH<7	P201 P209 P210 P211 P211B P219	
Hypoxic ischemic encephalopathy grade 2-3	P910 P911 P912 P913 P914 P915 P916 P916B P916C P916X	
Intracranial haemorrhage	P100 P101 P102 P103 P104 P108 P109 P520 P521 P523 P524 P525 P526 P528 P529	
Neonatal convulsions	P909 P909A P909B P909C	
Respiratory distress and meconium aspiration syndrome	P240 P241 P242 P248 P249 P220 P228 P229	
Invasive mechanical ventilation	DG021 DG022 DG002 DG026	
Cardiorespiratory resuscitation (intubation, ventilation, heart compressions)	DG017 DG018 DM004 DG010 DF017 DF012 DF028 + MBR-variables	ACIDOS, INTUB, HJMASS
Therapeutic hypothermia	DV034	
Extremely low birthweight. <1500 grams		BVIKT
Extremely preterm, <32 weeks of gestation		GRVBS
APGAR score <4 at five minutes		APGAR5
Congenital pneumonia	P230 P231 P232 P233 P234 P235 P236 P238 P239	
Congenital sepsis	P360 P361 P362 P363 P364 P365 P368 P369	
Birth trauma (fractures, neurological injury, retinal hemorrhage or facial nerve palsy, pulmonary hemorrhage, pneumothorax)	P110 P111 P112 P113 P114 P115 P119 P130 P131 P132 P133 P134 P138 P139 P150 P151 P152 P153 P154 P155 P156 P158 P159 P260 P261 P268 P269 P251 P252 P253	
Obstetric brachial plexus injury	P140 P141 P142 P143 P148 P149	
Hypoglycemia <2.2 mmol/l	P704A P704B	

Notes: Neonatal morbidity events are documented in the Swedish Medical Birth Register (MFR) and defined using medical diagnoses classified according to the International Classification of Diseases 10 (ICD-10) and procedures (using KVÅ-codes) during pregnancy and delivery. The morbidity events listed in this table are defined by the ICD-10 and KVÅ codes presented in column 2 and/or by the MFR-variables presented in column 3.

Table A3: Summary Statistics by IVF Status – Adverse Maternal Outcomes

	IVF		Non-IVF		Difference
	Mean	Std. Dev.	Mean	Std. Dev	(p-values)
Hemorrhage 1000ml	8.585	27.897	4.899	21.575	0.000
Hysterectomy	0.034	1.693	0.022	1.474	0.352
Post-birth surgery	0.304	5.479	0.089	2.975	0.057
Thromboembolism	0.046	2.142	0.029	1.699	0.221
Sepsis	0.568	7.456	0.310	5.559	0.000
Maternal deaths	0.000	0.000	0.004	0.660	0.565
Perineal lacerations 3-4	3.131	17.417	2.018	14.063	0.000
Anesthesia complications	0.011	1.071	0.023	1.501	0.737
Injury	0.126	3.550	0.086	2.925	0.573
Depression	0.034	1.855	0.046	2.136	0.000
Placenta complications	0.826	9.018	0.196	4.416	0.000
Multiple birth comp.	0.270	5.101	0.026	1.607	0.000
Ruptures	0.281	5.212	0.114	3.366	0.000
Eclampsia	0.103	3.211	0.056	2.364	0.016
Preeclampsia	7.145	25.704	3.580	18.574	0.000
Diabetes	0.981	9.811	0.794	8.875	0.588
Cervical Lacerations	0.528	7.245	0.289	5.366	0.000
Choriomenoitis	0.573	7.551	0.182	4.265	0.000
Wound infection	0.281	5.267	0.150	3.864	0.000
Endometriosis	0.568	7.456	0.309	5.551	0.000
Urinary Tract Infection	0.757	8.635	0.372	6.082	0.000

Notes: Summary statistics for specific maternal morbidity measures used to construct the maternal morbidity index are displayed separately by a mother's IVF status. Each measure is a binary indicator of whether the mother suffered a particular event, multiplied by 100 for ease of visualisation. p-values of tests for equality of means by group are reported in the final column.

Table A4: Summary Statistics by IVF Status – Adverse Child Outcomes

	IVF		Non-IVF		Difference
	Mean	Std. Dev.	Mean	Std. Dev	(p-values)
Umbilical chord pH<7	967.511	9789.098	588.402	7648.145	0.000
Hypoxic ischemic encephalopathy	59.723	2443.243	36.377	1906.918	0.653
Intracranial hematoma	167.224	4086.128	94.484	3072.384	0.174
Neonatal convulsions	226.947	4758.770	161.096	4010.447	0.372
Meconium aspiration	4753.942	21280.233	1841.502	13444.684	0.000
Mechanical ventilation	0.000	0.000	0.000	0.000	0.028
Cardiorespiratory resuscitation	692.785	8294.987	376.521	6124.572	0.001
Therapeutic hypothermia	0.000	0.000	0.000	0.000	0.720
APGAR < 4	474.337	6871.278	262.887	5120.514	0.002
Pneumonia	358.337	5975.750	321.247	5658.763	0.180
Sepsis	1457.238	11984.050	821.306	9025.313	0.000
Birth trauma	1015.289	10025.473	961.616	9758.949	0.329
Hypoglycemia	6473.961	24608.052	2757.293	16374.592	0.000
Plexus injury	95.557	3089.931	252.510	5018.692	0.014
Still birth	334.448	5773.815	327.861	5716.528	0.813
Extremely low birth weight	2962.255	16955.380	712.885	8413.114	0.000
Extremely Preterm	1158.624	10702.040	250.856	5002.275	0.000

Notes: Summary statistics for specific infant morbidity measures used to construct the neonatal morbidity index are displayed separately by their mother's IVF status. Each measure is a binary indicator of whether the child suffered a particular event, multiplied by 100,000 for ease of presentation. p-values of tests for equality of means by group are reported in the final column.

Table A5: Summary Statistics by IVF Status – Parental Characteristics

	IVF		Non-IVF		Difference
	Mean	Std. Dev.	Mean	Std. Dev.	(p-values)
Panel A: Principal Outcome Measures (Parents)					
Twin birth	0.19	0.39	0.01	0.11	0.000
N Births within 9 yrs	0.58	0.71	0.69	0.77	0.000
No future births	0.54	0.50	0.47	0.50	0.000
Maternal morbidity	0.23	0.42	0.13	0.34	0.000
Hospital nights, childbirth	5.21	8.20	3.58	6.29	0.000
Emergency C-section	0.17	0.37	0.08	0.28	0.000
Wage earnings, mother	192983	138751	162310	119960	0.000
Sickness benefits, mother	8126.6	14067.7	6823.3	12997.1	0.000
Wage earnings, father	400782	361257	337078	262925	0.000
Sickness benefits, father	4481.3	14636.1	4412.1	13724.5	0.000
Panel B: Covariates					
Age, mother	32.37	3.60	29.24	4.57	0.000
BMI, mother	24.53	4.11	24.38	4.31	0.346
Born in Sweden, mother	0.87	0.34	0.83	0.38	0.000
First time mother	0.74	0.44	0.44	0.50	0.000
Smoking, mother	0.06	0.24	0.11	0.32	0.000
Elementary, mother	0.08	0.27	0.13	0.34	0.000
High school, mother	0.49	0.50	0.51	0.50	0.020
University, mother	0.43	0.50	0.36	0.48	0.000
Age, father	35.58	5.23	32.66	5.73	0.000
Born in Sweden, father	0.87	0.34	0.82	0.39	0.000
Elementary, father	0.10	0.30	0.14	0.34	0.000
High school, father	0.35	0.48	0.34	0.47	0.000
University, father	0.55	0.50	0.52	0.50	0.000
Total Observations	17,565		877,775		

Notes: Characteristics and outcomes of IVF and non-IVF parents from the Swedish Medical Birth Registry and LISA Registry are displayed, along with p-values testing for equality across groups. Total observations indicated at the foot of the table refer to the total number of mothers aged under 39, and hence used in principal models. In a number of robustness checks, we include additionally women over the age of 39, resulting in a sample of 19,964 IVF mothers and 921,287 non-IVF mothers.

Table A6: Summary Statistics by IVF Status – Child Outcomes

	IVF		Non-IVF		Difference
	Mean	Std. Dev.	Mean	Std. Dev.	(p-values)
Principal Outcome Measures (Children)					
Birth weight	3182.382	760.151	3546.027	591.518	0.000
Gestational age (weeks)	38.269	2.923	39.335	1.968	0.000
Hospital nights, age 0	4.729	16.392	1.650	9.080	0.000
Hospital nights, ages 1-8	1.338	8.781	1.057	7.718	0.000
AAR score < 7	0.021	0.143	0.011	0.107	0.000
Severe neonatal morbidity	0.148	0.355	0.074	0.262	0.000
Infant mortality	0.009	0.094	0.006	0.075	0.000
Under 5 mortality	0.002	0.046	0.001	0.033	0.000
Total Observations	19,563		888,675		

Notes: Characteristics of children born as a result of IVF procedures and non-IVF procedures are displayed. All measures are generated from microdata in the Swedish Medical Birth Registry, and p-values testing for equality across groups are reported. Total observations indicated at the foot of the table refer to the total number of children born to mothers aged under 39 (which is larger than the total number of mothers, given multiple births). In a number of robustness checks, we extend to include additionally women over the age of 39, resulting in a sample of 22,183 IVF births and 932,822 non-IVF births.

Table A7: Summary Statistics of Singleton and Twin Births by IVF status

	Twins			Singletons		
	IVF mean	Non-IVF mean	p-values	IVF mean	Non-IVF mean	p-values
Principal Outcome Measures (Children)						
Birth weight	2559.1	2569.4	0.342	3470.5	3557.6	0.000
Gestational age (weeks)	36.1	36.1	0.757	39.2	39.4	0.000
Hospital nights, age 0	9.13	8.47	0.067	2.19	1.41	0.000
Hospital nights, ages 1-8	1.55	1.32	0.183	1.11	0.992	0.054
APGAR score < 7	0.032	0.033	0.804	0.015	0.011	0.000
Severe neonatal morbidity	0.248	0.226	0.002	0.094	0.067	0.000
Infant mortality	0.015	0.016	0.458	0.006	0.005	0.159
Under 5 mortality	0.002	0.002	0.994	0.001	0.001	0.116

Notes: Mean values of indicators of health at birth are presented for twins (left-hand panel) and singleton births (right-hand panel). Identical measures as those used in principal models of the paper are displayed. Tests of equality are presented between IVF and non-IVF births within each group of singletons or twins (p-values corresponding to these tests are reported).

Table A8: Proportion of IVF births by group

Age		Birth Order		Education		BMI	
Group	Prop.	Group	Prop.	Group	Prop.	Group	Prop.
Age <25	0.002	1st pregnancy	0.032	Elementary	0.010	Normal weight	0.020
Age 25-29	0.010	2nd pregnancy	0.012	High school	0.019	Overweight	0.022
Age 30-34	0.025	3rd or higher	0.004	University	0.025	Obese	0.021
Age 35-38	0.044	–	–	–	–	–	–
Age >38	0.052	–	–	–	–	–	–

Notes: Each panel displays the proportion of all births among mothers who meet specific age, birth order, education or BMI criteria noted in the right hand panel. All proportions refer to all live births registered in the Medical Birth Registry over the period under study (1998-2016).

Table A9: The Impact of SET on Neonatal Morbidity

Panel A	APGAR < 4	Pneumonia	Sepsis	Birth Trauma	Hypogly- cemia	Plexus Injury
SET reform	-0.001	0.001	-0.001	0.000	-0.019***	0.001**
	[0.001]	[0.001]	[0.002]	[0.001]	[0.004]	[0.000]
IVF	0.002*	-0.000	0.004***	-0.001	0.029***	-0.002***
	[0.001]	[0.001]	[0.001]	[0.001]	[0.003]	[0.000]
Scaled Impact of SET	-0.335	-1.116	-0.231	-0.336	-0.646	-0.684
Observations	900,261	908,232	908,232	908,232	908,232	908,232
Mean of Dep. Var. IVF	0.005	0.004	0.015	0.010	0.065	0.001
Mean of Dep. Var. Full	0.003	0.003	0.008	0.010	0.028	0.002
Uncorrected p-value (SET)	0.615	0.553	0.637	0.904	0.000	0.032
Corrected p-value (SET)	1	1	1	1	< 0.01	0.378
Panel B	Still- birth	Extremely Low BW	Extremely Preterm	pH< 7	Hypo. Isch. Encephalop.	Intra. Haem.
SET reform	0.002*	-0.013***	-0.005***	-0.002*	-0.000	0.001
	[0.001]	[0.003]	[0.002]	[0.001]	[0.000]	[0.001]
IVF	-0.000	0.021***	0.009***	0.001	0.000	0.000
	[0.001]	[0.002]	[0.002]	[0.001]	[0.000]	[0.000]
Scaled Impact of SET	-7.356	-0.631	-0.631	-1.634	-7.491	2.944
Observations	908,232	908,232	908,232	908,232	908,232	908,232
Mean of Dep. Var. IVF	0.003	0.030	0.012	0.010	0.001	0.002
Mean of Dep. Var. Full	0.003	0.008	0.003	0.006	0.000	0.001
Uncorrected p-value (SET)	0.064	0.000	0.002	0.086	0.878	0.153
Corrected p-value (SET)	0.701	< 0.01	0.023	0.856	1	1
Panel C	Neo. Convul.	Meconium Aspiration	Mechanical Ventilation	Cardioresp.	Ther Hypothermia	
SET reform	-0.001	-0.010***	0.000	-0.002	-0.000***	
	[0.001]	[0.003]	[0.000]	[0.001]	[0.000]	
IVF	0.000	0.024***	-0.000	0.002	-0.000*	
	[0.001]	[0.003]	[0.000]	[0.001]	[0.000]	
Scaled Impact of SET	-7.170	-0.407	-9.032	-1.049	2.339	
Observations	908,232	908,232	908,232	908,232	908,232	
Mean of Dep. Var. IVF	0.002	0.048	0.000	0.007	0.000	
Mean of Dep. Var. Full	0.002	0.019	0.000	0.004	0.000	
Uncorrected p-value (SET)	0.274	0.004	0.267	0.188	0.001	
Corrected p-value (SET)	1	0.046	1	1	0.019	

Notes: Each column presents a separate two-way FE estimate of the impact of the SET reform on neonatal health and morbidity measures following specification 2. Estimated date of conception fixed effects are included in all regressions. Standard errors are clustered by mother. The corrected p-value refers to p-values based on Holm's FWER correction considering all 17 outcomes in this Table. * p<0.1, ** p<0.05, *** p<0.01.

Table A10: The Impact of SET on Maternal Morbidity

Panel A	Hameorr-hage	Hyster-ectomy	Post-birth surgery	Thrombo-embolism	Sepsis	Maternal Death	Perineal Lacer-ation
SET reform	0.034	0.052	0.040	-0.018	-0.090	0.009	-0.052
	[0.456]	[0.036]	[0.098]	[0.028]	[0.113]	[0.010]	[0.290]
IVF	2.808***	0.008	0.209***	0.008	0.183**	-0.007***	-0.171
	[0.349]	[0.022]	[0.073]	[0.025]	[0.092]	[0.002]	[0.228]
Scaled Impact of SET	0.012	6.808	0.194	-2.187	-0.490	-1.362	0.306
Observations	895,336	895,336	895,336	895,336	895,336	895,336	895,336
Mean of Dep. Var. IVF	9.394	0.036	0.370	0.043	0.605	0.000	3.743
Mean of Dep. Var. Full	5.342	0.025	0.118	0.032	0.323	0.004	2.640
Uncorrected p-value (SET)	0.941	0.145	0.678	0.529	0.427	0.342	0.856
Corrected p-value (SET)	1	1	1	1	1	1	1

Panel B	Anastesia complic-ations	Injury	Depre-ssion	Placenta complic-ations	Mult. Birth Comp.	Ruptures	Eclamp-sia
SET reform	-0.005	-0.057	-0.041	0.165	-0.160**	-0.056	-0.020
	[0.015]	[0.055]	[0.042]	[0.134]	[0.074]	[0.081]	[0.054]
IVF	-0.020	0.042	-0.029	0.453***	0.268***	0.125*	0.036
	[0.015]	[0.046]	[0.025]	[0.099]	[0.065]	[0.065]	[0.043]
Scaled Impact of SET	0.243	-1.342	1.406	0.364	-0.596	-0.451	-0.557
Observations	895,336	895,336	895,336	895,336	895,336	895,336	895,336
Mean of Dep. Var. IVF	0.014	0.142	0.043	0.690	0.306	0.306	0.128
Mean of Dep. Var. Full	0.026	0.093	0.053	0.212	0.033	0.128	0.063
Uncorrected p-value (SET)	0.749	0.307	0.323	0.218	0.032	0.485	0.711
Corrected p-value (SET)	1	1	1	1	0.664	1	1

Panel C	Pre-Ec-lampsia	Diabetes	Cervical Lacerat.	Choriomen-oitis	Wound Infection	Endometri	UTI
SET reform	-0.787**	0.153	-0.047	-0.058	0.001	-0.089	-0.160
	[0.384]	[0.170]	[0.113]	[0.116]	[0.078]	[0.113]	[0.132]
IVF	2.011***	0.024	0.189**	0.308***	0.028	0.185**	0.291***
	[0.308]	[0.127]	[0.089]	[0.092]	[0.063]	[0.092]	[0.109]
Scaled Impact of SET	-0.391	6.376	-0.248	-0.188	0.023	-0.479	-0.551
Observations	895,336	895,336	895,336	895,336	895,336	895,336	895,336
Mean of Dep. Var. IVF	7.180	1.110	0.555	0.598	0.278	0.605	0.840
Mean of Dep. Var. Full	3.742	0.852	0.298	0.205	0.170	0.323	0.414
Uncorrected p-value (SET)	0.040	0.367	0.678	0.616	0.993	0.432	0.225
Corrected p-value (SET)	0.804	1	1	1	1	1	1

Notes: Each column presents a separate two-way FE estimate of the impact of the SET reform on maternal morbidity outcomes following specification 2. Estimated date of conception fixed effects are included in all regressions. Standard errors are clustered by mother. The corrected p-value refers to p-values based on Holm's FWER correction considering all 21 outcomes in this Table. * p<0.1, ** p<0.05, *** p<0.01.

Table A11: The impact of SET on extensive vs intensive margin earnings

	Mothers		Fathers		
	Non-zero wage	Wage income (intensive margin)	Non-zero wage	Wage income (intensive margin)	Relative Female Wage
SET reform	0.005**	10097.019***	0.002	-2822.879	12889.284**
	(0.002)	(2066.624)	(0.003)	(5103.602)	(5311.522)
IVF	-0.004*	-1494.855	-0.001	23378.530***	-24163.280***
	(0.002)	(1535.096)	(0.002)	(4212.426)	(4339.532)
Scaled Impact of SET	-1.363	-6.755	-2.590	-0.121	-0.533
Observations	893,747	855,717	893,793	863,761	892,718
Mean of Dep. Var. IVF	0.970	198851.041	0.970	413089.735	-207676.944
Mean of Dep. Var. Full	0.958	169996.115	0.966	349960.027	-175243.521
Uncorrected p-value (SET)	0.038	0.000	0.530	0.580	0.015
Corrected p-value (SET)	0.342	< 0.001	1.00	1.00	0.165

Notes: Each column presents a separate two-way FE estimate of the impact of the SET reform on maternal or paternal labour market participation (measured as reporting at least some wage income in the years following birth), and wage income at the extensive margin (wage among all individuals reporting non-zero wages). The final column estimates the impact of the SET reform on the relative female wage to the male wage within all households. Wage earning refer to averages over the 9 years following each mother or father's birth, and are generated by following parents up to (a maximum of) 2016 in the LISA register. Estimated date of conception fixed effects are included in all regressions. The scaled impact of SET refers to the proportional impact of the SET reform compared to the difference between IVF and non-IVF mothers. This is the ratio of the coefficient in row 1 to the coefficient in row 2. Corrected p-values are calculated using Holm's Family Wise Error Rate correction across all models estimated in Tables 1-4. Standard errors are clustered by mother. * p<0.1, ** p<0.05, *** p<0.01.

Table A12: The SET reform and maternal and paternal characteristics

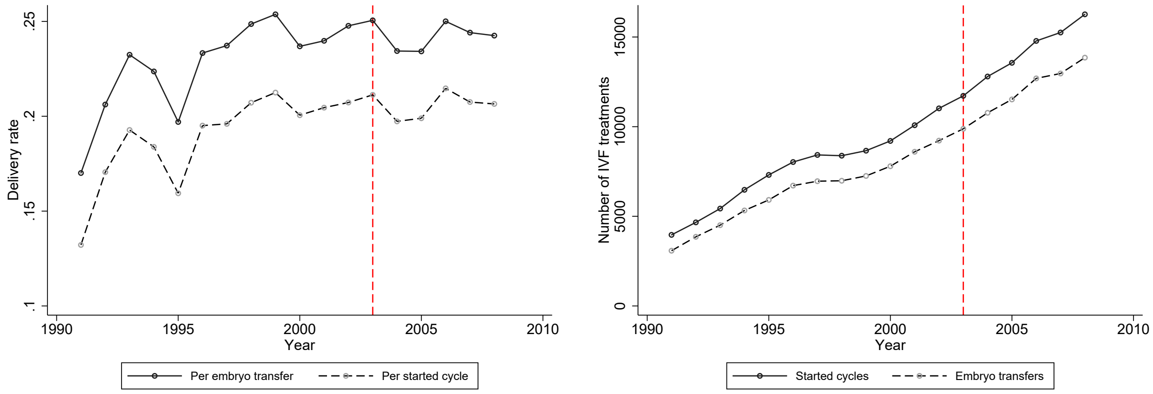
	Age (1)	Primary (2)	Secondary (3)	Tertiary (4)	Native (5)	BMI (6)	Smoke (7)
Panel A: Mothers							
SET reform	0.029	-0.008**	-0.010	0.018**	0.000	-0.090	0.005
	[0.056]	[0.004]	[0.008]	[0.008]	[0.005]	[0.068]	[0.003]
IVF	3.124***	-0.055***	-0.016***	0.071***	0.044***	0.148***	-0.053***
	[0.044]	[0.003]	[0.006]	[0.006]	[0.004]	[0.054]	[0.003]
Proportional Change	0.001	-0.105	-0.021	0.042	0.000	-0.004	0.074
Observations	895,340	860,550	860,550	860,550	894,999	783,673	841,292
Mean of Dep. Var. IVF	32.366	0.077	0.489	0.434	0.869	24.534	0.061
Mean of Dep. Var. Full	29.287	0.131	0.506	0.363	0.828	24.384	0.112
Uncorrected p-value (SET)	0.609	0.040	0.193	0.020	0.942	0.190	0.192
Corrected p-value (SET)	1.00	0.399	1.00	0.220	1.00	1.00	1.00
Panel B: Fathers							
SET reform	0.059	-0.009*	0.013*	-0.005	0.017***	–	–
	[0.082]	[0.005]	[0.007]	[0.008]	[0.005]	–	–
IVF	2.921***	-0.038***	0.015**	0.023***	0.053***	–	–
	[0.064]	[0.004]	[0.006]	[0.006]	[0.004]	–	–
Proportional Change	0.002	-0.086	0.038	-0.009	0.019	–	–
Observations	895,340	876,923	876,923	876,923	895,340	–	–
Mean of Dep. Var. IVF	35.577	0.099	0.351	0.550	0.869	–	–
Mean of Dep. Var. Full	32.706	0.137	0.339	0.524	0.819	–	–
Uncorrected p-value (SET)	0.474	0.060	0.065	0.534	0.002	–	–
Corrected p-value (SET)	1.00	0.539	0.539	1.00	0.024	–	–

Notes: Each column presents a separate two-way FE regression where parental characteristics are regressed on the SET reform and IVF indicator following specification 2. Models replicate those from Table 1, however replacing outcome variables with observable parental characteristics (control variables included in all baseline models in the paper). Maternal characteristics are presented in Panel A, with similar paternal characteristics in Panel B. In the case of BMI and a woman's smoking status during pregnancy, these are only observed for mothers. Estimated date of conception fixed effects are included in all regressions. "Proportional Change" in panel footers refers to the parameter estimated on SET reform divided by the baseline dependent variable mean for IVF users. Standard errors are clustered by mother. Corrected p-value refers to p-values from Holm's multiple hypothesis correction based on the hypotheses tested in this Table. * p<0.1, ** p<0.05, *** p<0.01.

Figure A1: Trends in delivery rates and IVF treatments

(a) Deliveries per embryo transfer and cycle

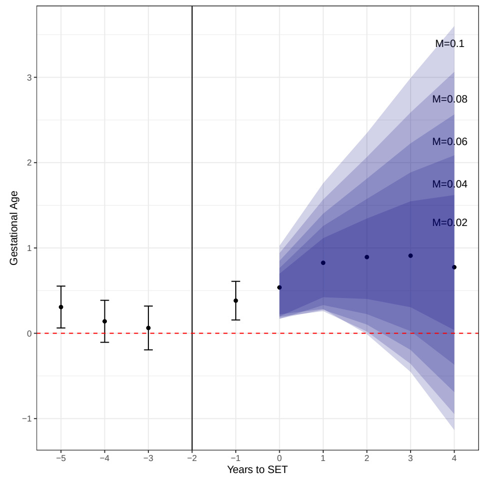
(b) Number of IVF treatments



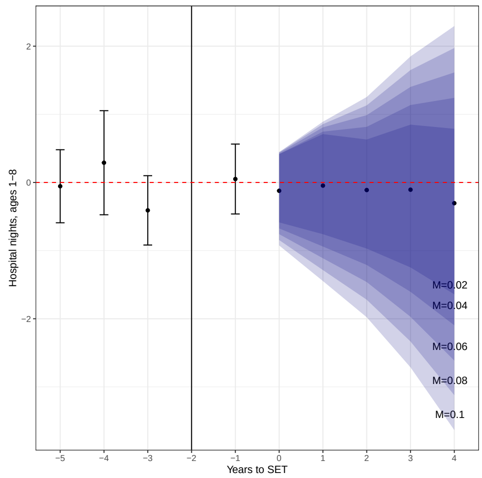
Notes: Annual trends in deliveries per transfer/cycle and the number of IVF treatments are based on aggregate data collected from annual reports by the Swedish National Board of Health and Welfare and presented in Figures A1a and A1b. Started cycles refer to all IVF procedures initiated, whether or not an embryo was eventually transferred, while embryo transfers refers only to those procedures in which an embryo was transferred to the woman. The red vertical line indicates the year of the SET reform.

Figure A2: Honest DiD Bounds – Alternative Child Health Measures

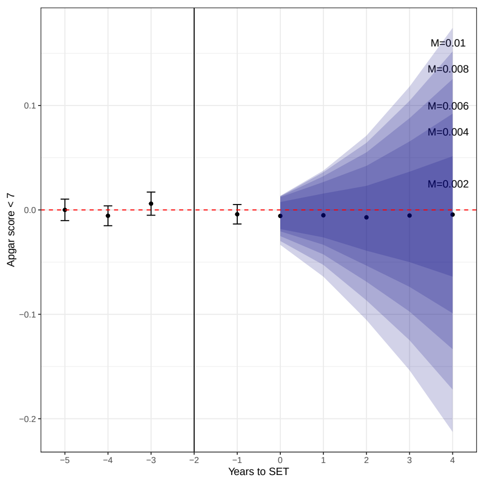
(a) Gestional Age (weeks)



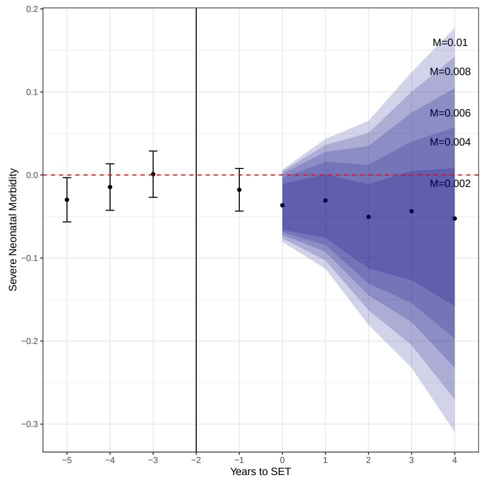
(b) Hospital nights, ages 1-8



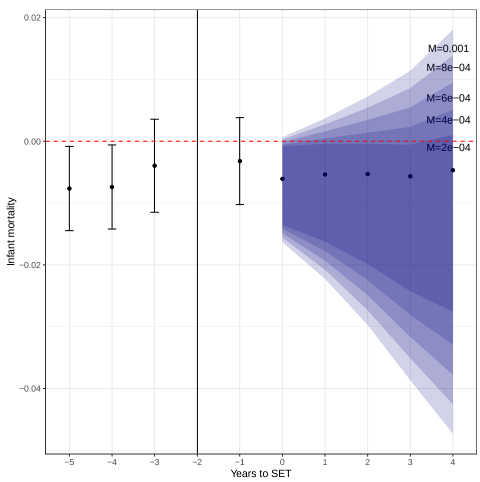
(c) APGAR score < 7



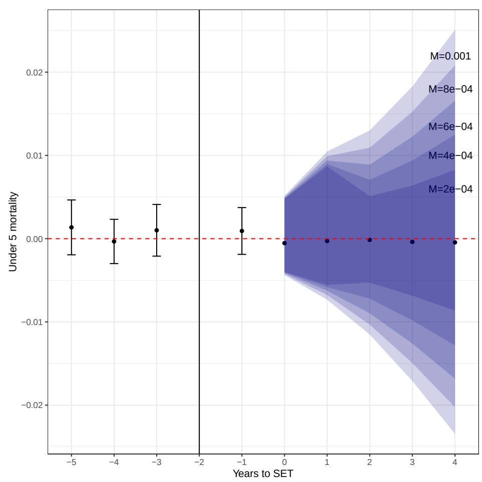
(d) Severe Neonatal Morbidity



(e) Infant mortality



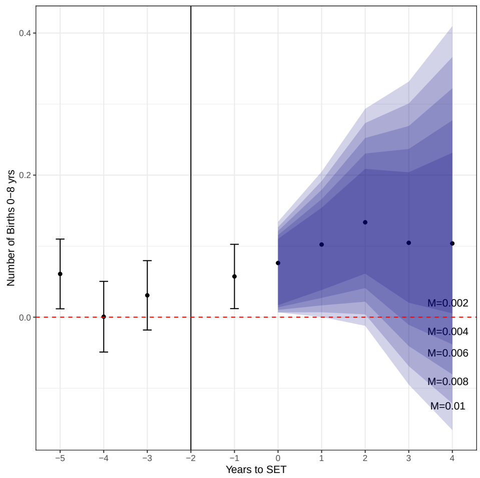
(f) Under 5 mortality



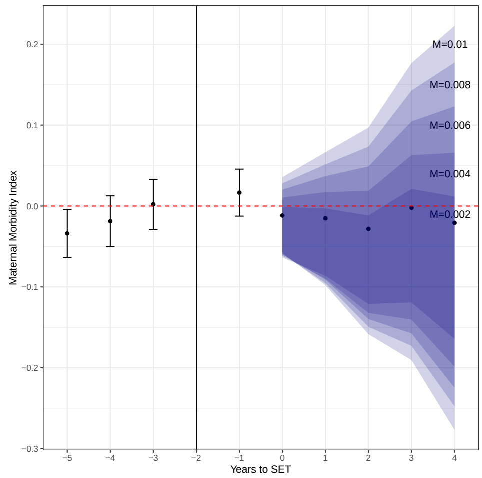
Notes: Refer to notes to Figure 8. Identical "Honest Difference-in-Differences" procedures are implemented, here focusing on child health measures not presented in Figure 8.

Figure A3: Honest DiD Bounds – Alternative Parental Outcomes

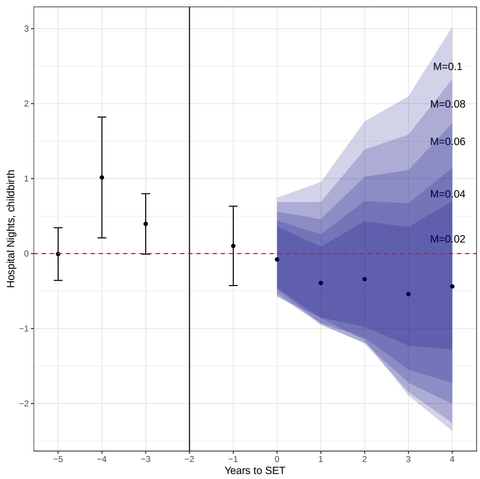
(a) Number of future births within 9 years



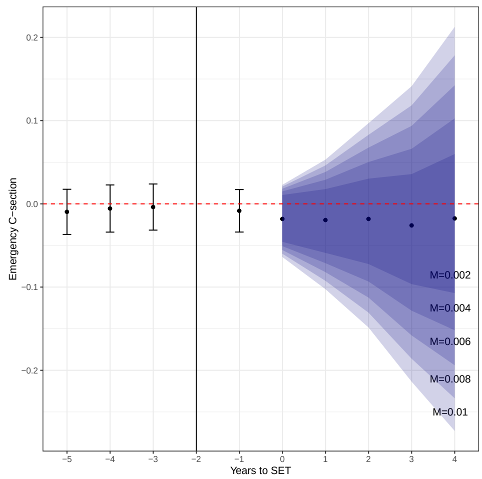
(b) Maternal Morbidity Index



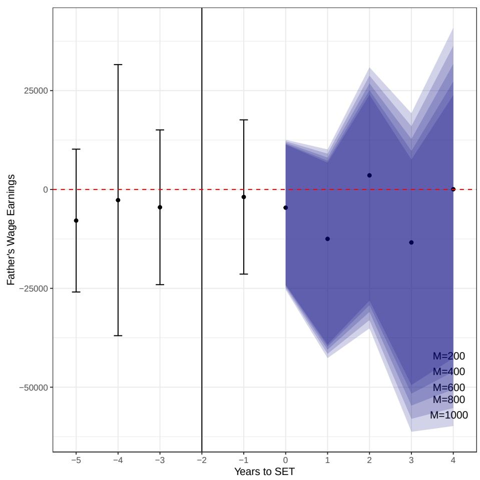
(c) Hospital nights



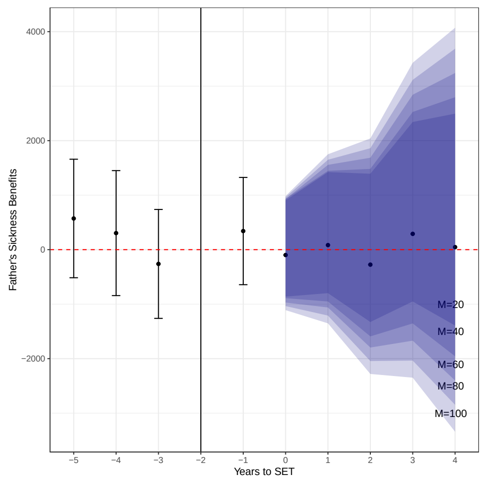
(d) Emergency C-section



(e) Wage earnings, father



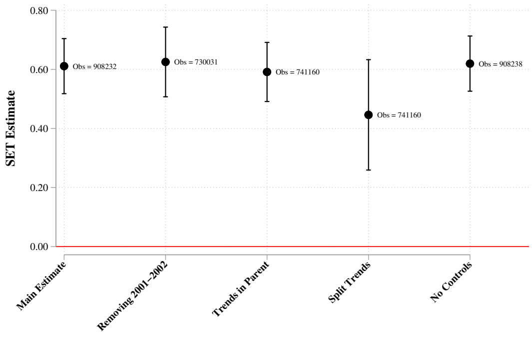
(f) Income from sickness leave, father



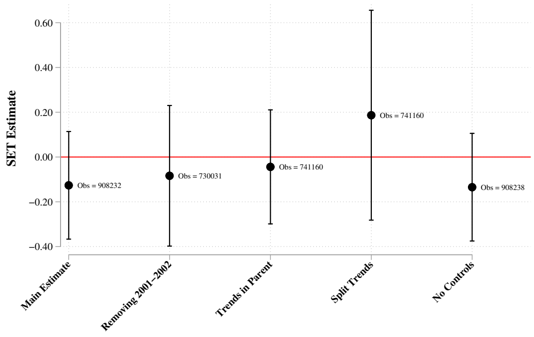
Notes: Refer to notes to Figure 8. Identical "Honest Difference-in-Differences" procedures are implemented, here focusing on additional parental outcomes (fertility, health, and labour market measures) not presented in Figure 8.

Figure A4: Alternative Models and Samples – Additional Child Health Measures

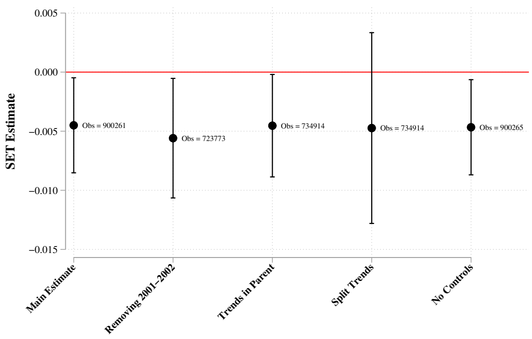
(a) Gestational Age (weeks)



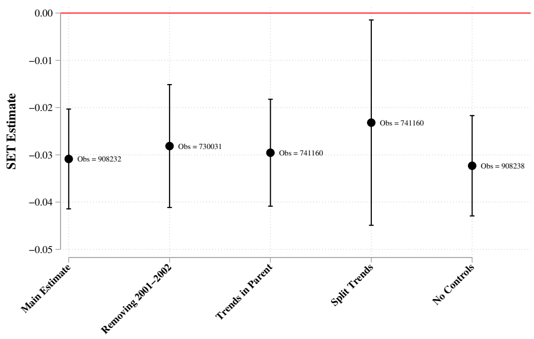
(b) Hospital nights, ages 1–8



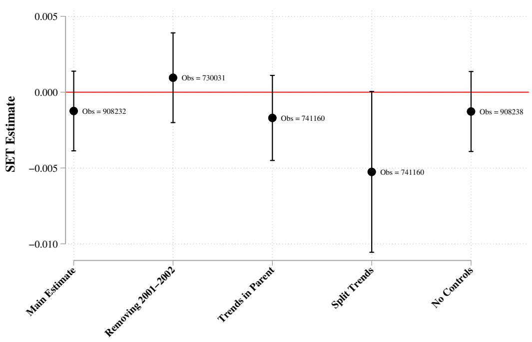
(c) APGAR score < 7



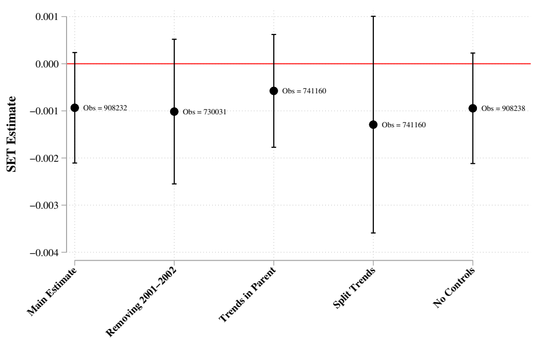
(d) Severe Neonatal Morbidity



(e) Infant Mortality



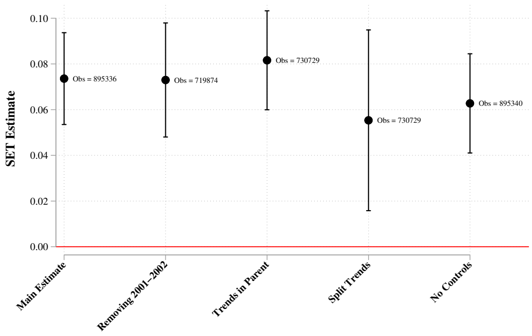
(f) Under 5 mortality



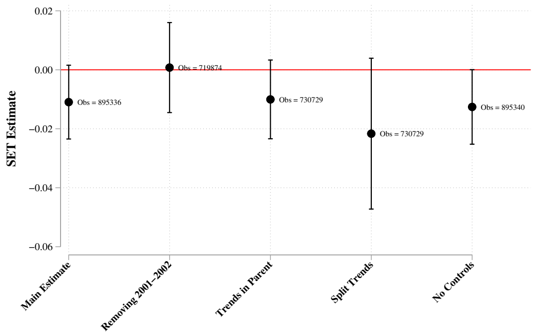
Notes: Refer to notes to Figure 9. Identical robustness plots are displayed, here focusing on additional child health measures not presented in Figure 9.

Figure A5: Alternative Models and Samples – Additional Parental Outcomes

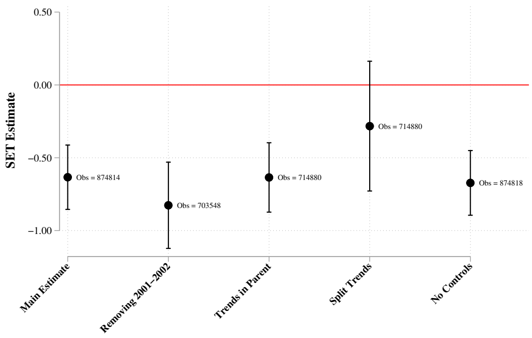
(a) Number of future births within 9 yrs



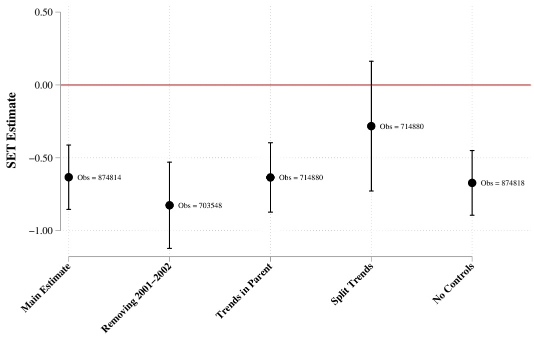
(b) Maternal Morbidity Index



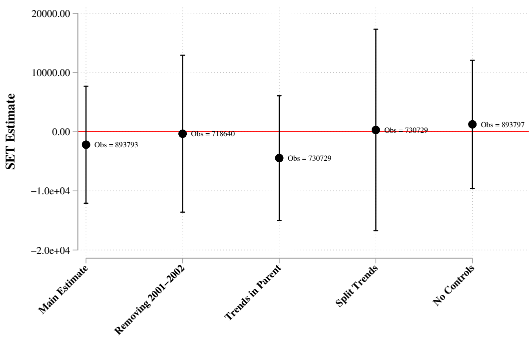
(c) Hospital nights



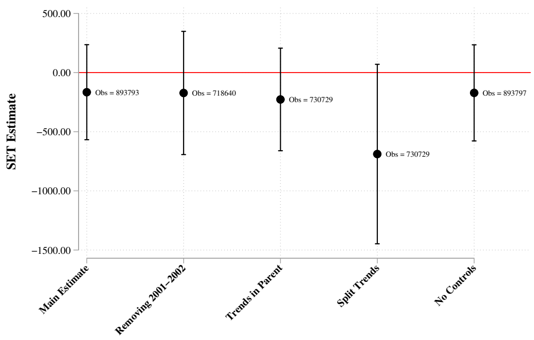
(d) Emergency C-section



(e) Wage earnings, father



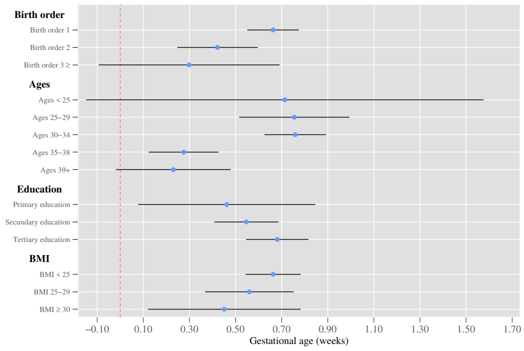
(f) Income from sickness leave, father



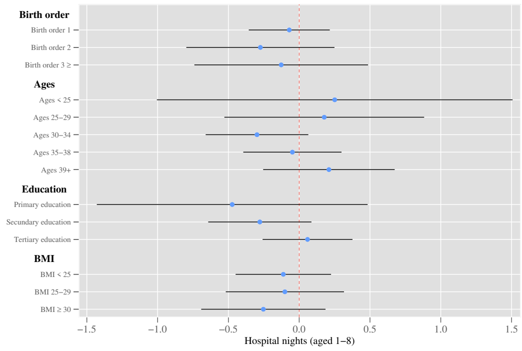
Notes: Refer to notes to Figure 9. Identical robustness plots are displayed, here focusing on additional parental outcomes (fertility, health, and labour market measures) not presented in Figure 9.

Figure A6: Reform Heterogeneity – Additional Child Health Measures

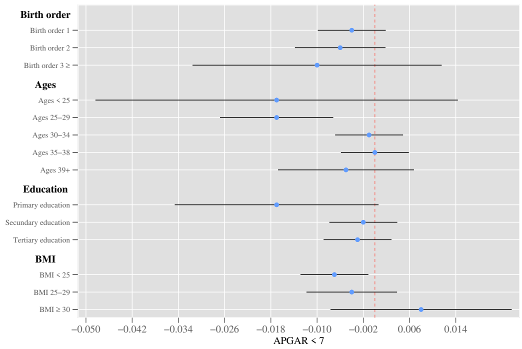
(a) Gestational Age (weeks)



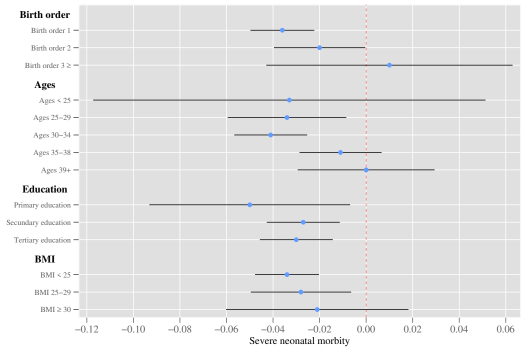
(b) Hospital nights, ages 1–8



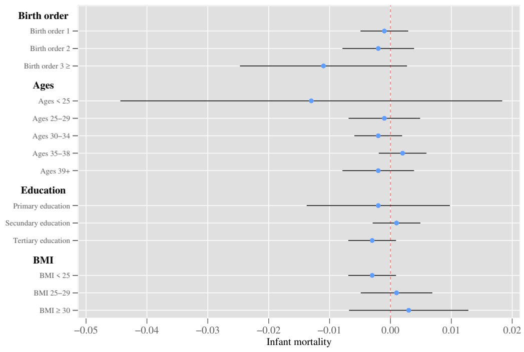
(c) APGAR score < 7



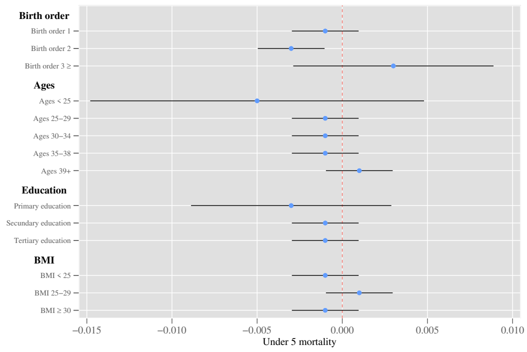
(d) Severe Neonatal Morbidity



(e) Infant Mortality



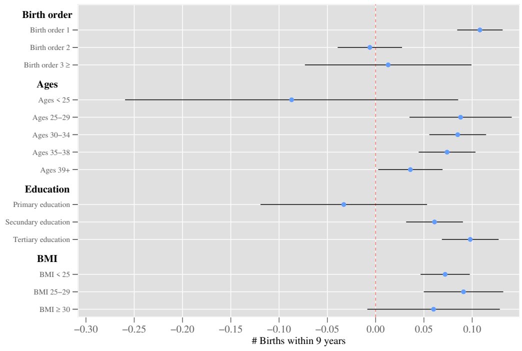
(f) Under 5 mortality



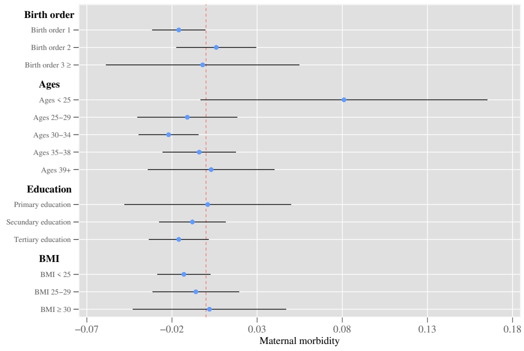
Notes: Refer to notes to Figure 10. Identical group-specific estimates are plotted for the groups indicated on vertical plot axes, however for alternative measures of child health not documented in Figure 10.

Figure A7: Reform Heterogeneity – Additional Parental Outcomes

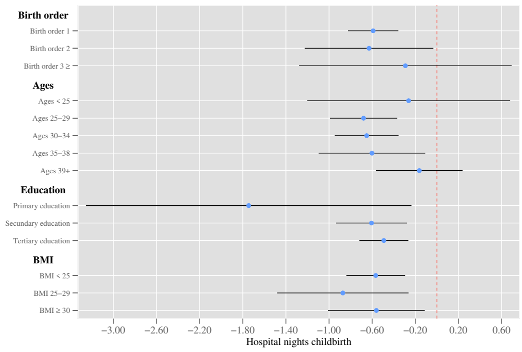
(a) Number of future births within 9 yrs



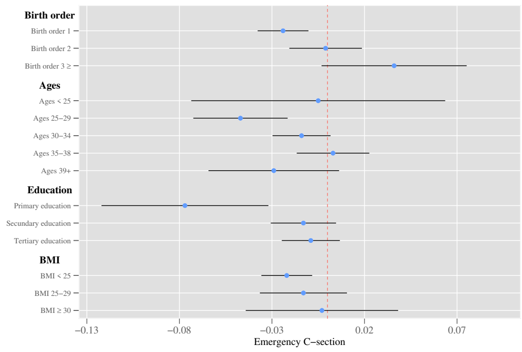
(b) Maternal Morbidity Index



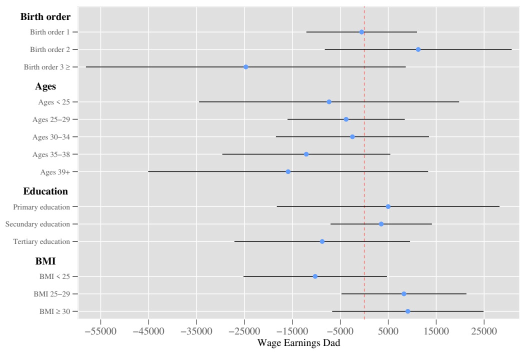
(c) Hospital nights



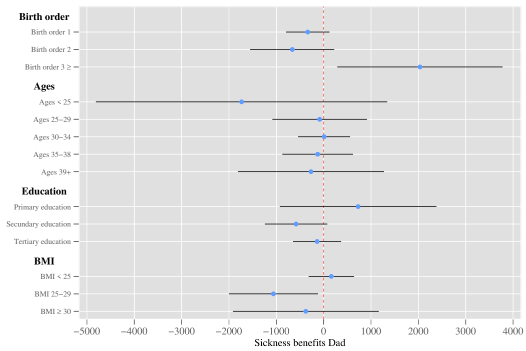
(d) Emergency C-section



(e) Wage earnings, father



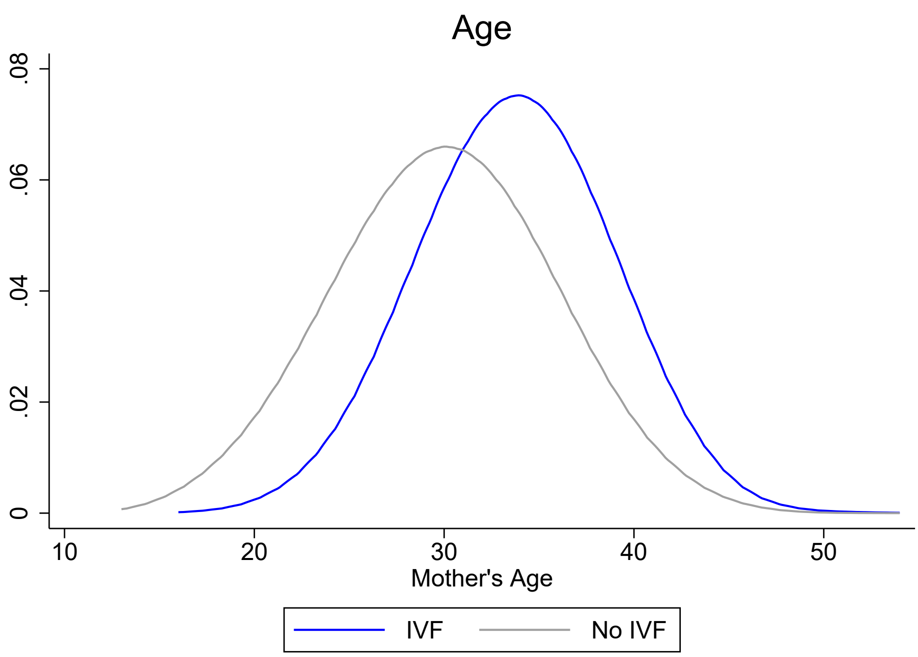
(f) Income from sickness leave, father



Notes: Refer to notes to Figure 10. Identical group-specific estimates are plotted for the groups indicated on vertical plot axes, however for alternative measures of child health not documented in Figure 10.

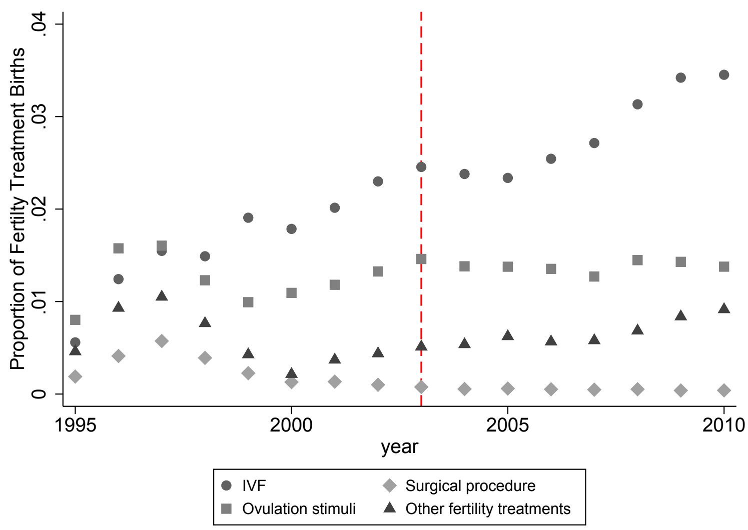
B Descriptive Plots and Trends in Principal Outcome Measures by IVF Status

Figure B1: Maternal Age



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

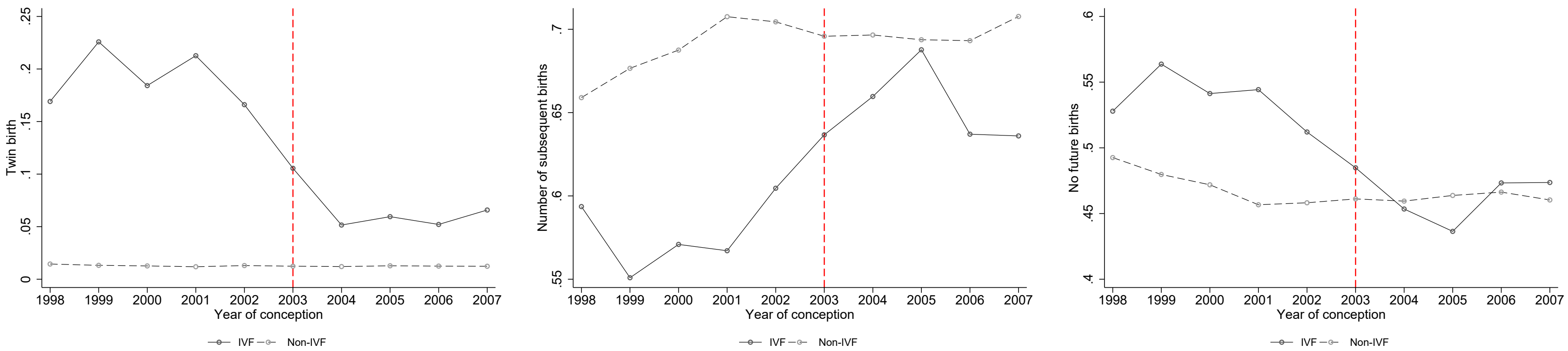
Figure B2: ART treatments



Notes: Data are obtained from the Swedish Medical Birth Registry, presenting trends in different classes of recorder ART treatments. The red-vertical line represents the year of the SET reform.

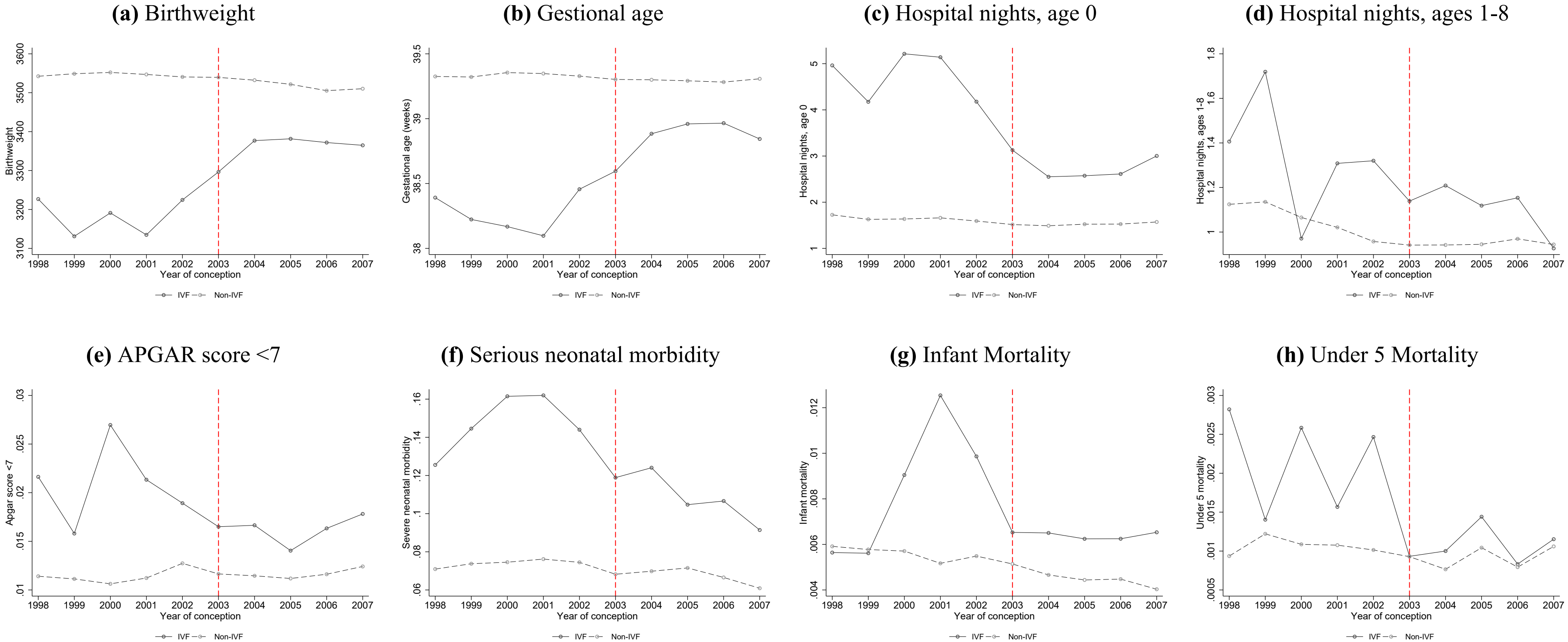
Figure B3: Trends in fertility outcomes

(a) Twin birth **(b)** Number of future births within 8 yrs **(c)** No more births



Notes: Summary plots are based on microdata obtained from the Swedish Medical Birth Registry covering the principal estimation sample from the paper, documenting raw averages in fertility outcomes by IVF status over time. The red-vertical line represents the year of the SET reform.

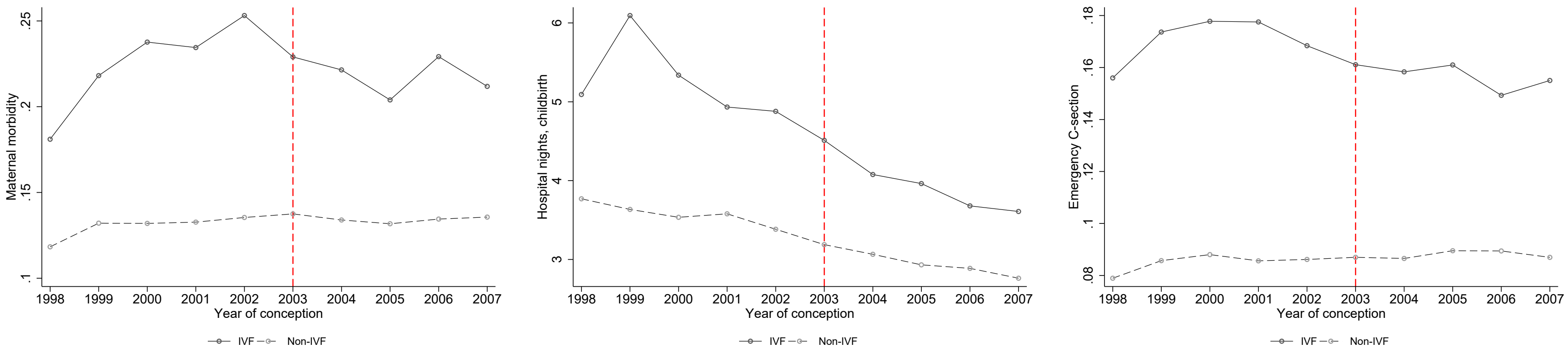
Figure B4: Trends neonatal and child health outcomes



Notes: Summary plots are based on microdata obtained from the Swedish Medical Birth Registry covering the principal estimation sample from the paper, documenting raw averages in outcomes by a child's conception status (IVF or non-IVF) over time. The red-vertical line represents the year of the SET reform.

Figure B5: Trends in maternal health measures

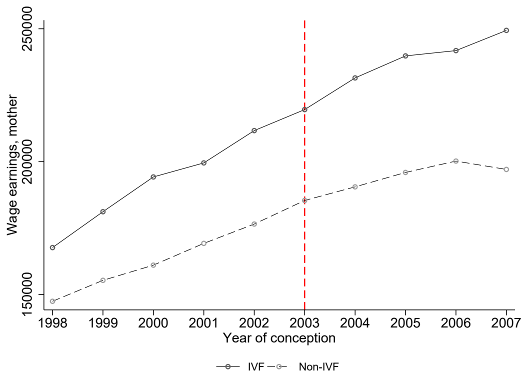
(a) Maternal Morbidity Index **(b)** Hospital nights **(c)** Emergency C-section



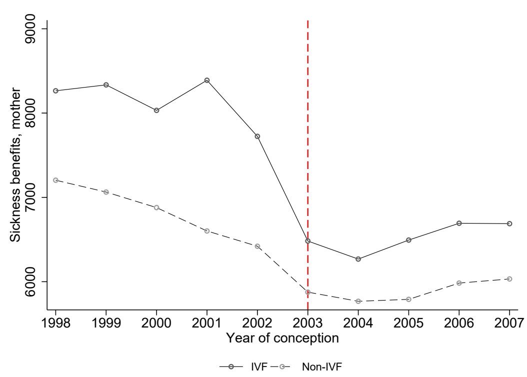
Notes: Summary plots are based on microdata obtained from the Swedish Medical Birth Registry and the Swedish National Patient Registry covering the principal estimation sample of the paper, documenting raw averages in outcomes by IVF status over time. The red-vertical line represents the year of the SET reform.

Figure B6: Trends in labour market outcomes

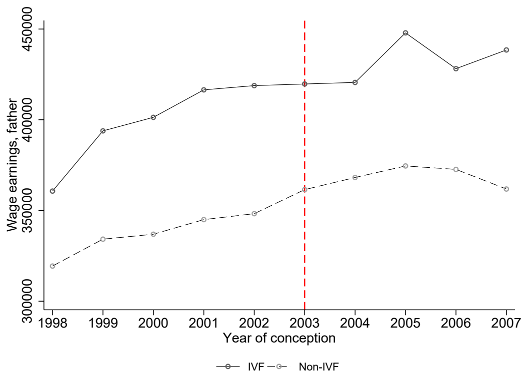
(a) Wage earnings, mothers



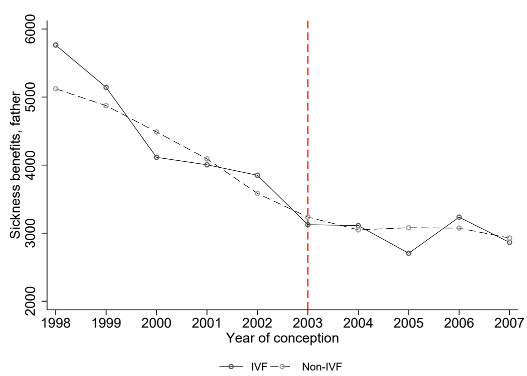
(b) Income from sickness benefits, mothers



(c) Wage earnings, fathers



(d) Income from sickness benefits, fathers



Notes: Summary plots are based on microdata obtained from the Longitudinal integration database for health insurance and labor market studies covering the principal estimation sample of the paper, documenting raw averages in outcomes by parental IVF status over time. The red-vertical line represents the year of the SET reform. Wage earnings and sickness benefits refer to averages in the 9 years following birth for all parents who give birth in a given year.

C Within Mother Variation and Identification

To observe why a standard mother fixed effect model does not have good properties when we are interested in estimating the parameter on a binary interaction term, consider the simplified model laid out below. We present this analysis in terms of first differences rather than mother fixed effects as it simplifies the underlying variation (all variation is binary), however these results hold in the case of mother fixed effects, as the identical underlying variation just needs to be considered in terms of deviations from the mean in each of the dependent variables, and scaled accordingly. The first differences model can be written:

$$(y_i - y_j) = \beta_0 + \beta_1(IVF_i - IVF_j) + \beta_2(IVF \cdot SET_i - IVF \cdot SET_j) + \boldsymbol{X}'\gamma + (\varepsilon_i - \varepsilon_j),$$

where y_i and y_j refer to outcomes following a mother's child i and j respectively, IVF_i and IVF_j refer to the mother's IVF status on each birth (1 if IVF, 0 if not), and $IVF \cdot SET$ refers to the binary IVF variable interacted with SET, a measure taking 1 if the birth is in the post-SET period, and 0 if it is in the pre-SET period. The vector $\boldsymbol{X}$ includes controls including year of birth fixed effects. To fix ideas below we will consider a mother with 2 births, and will consider y as her outcome in the years following the birth.

The parameter β_1 has a standard panel interpretation, which we discuss before passing on to the problematic parameter β_2. Formally, $\beta_1 \equiv \frac{\partial(y_i - y_j)}{\partial(IVF_i - IVF_j)}$, which in the case of binary variables can be simplified to[41]:

$$\beta_1 \equiv E[(y_i - y_j)|(IVF_i - IVF_j) = 1] - E[(y_i - y_j)|(IVF_i - IVF_j) = 0], \quad (3)$$

which is precisely what we wish to capture in the mother panel regression. In words, if y refers to the mother's salary, this is the salary change for mothers who have one IVF birth and one non-IVF birth (and hence for whom $IVF_i - IVF_j = 1$), compared to those who had two births of the same type (IVF or non-IVF), and hence for whom $(IVF_i - IVF_j) = 0$.

However, now consider the parameter β_2. This will similarly capture the mean differences in outcomes for mothers who have within birth variation in the dependent variable $(IVF \cdot SET_i - IVF \cdot SET_j)$, compared to mothers who do not have variation in this variable. To see that this is problematic as it includes a number of undesired comparisons, we lay out all the potential birth combinations based on IVF and SET below. Combinations in red are combinations for which $(IVF \cdot SET_i - IVF \cdot SET_j)$ equals zero, while combinations in green are those for which

[41] Strictly speaking, the below assumes that $IVF_i - IVF_j$ can only take the values of 0 or 1. In practice, it can also take the value of -1 (when $IVF_i = 0$ and $IVF_J = 1$). However, the model can always be re-written such that the IVF birth is listed first on both sides, in which case the restriction to values of 0 and 1 for the first difference just simplifies exposition without losing generality.

$(IVF \cdot SET_i - IVF \cdot SET_j)$ equals 1.[42]

	Birth 1	Birth 2
1.	IVF = 0, SET = 0	IVF = 0, SET = 0
2.	IVF = 0, SET = 0	IVF = 1, SET = 0
3.	IVF = 0, SET = 0	IVF = 0, SET = 1
4.	IVF = 0, SET = 0	IVF = 1, SET = 1
5.	IVF = 1, SET = 0	IVF = 0, SET = 0
6.	IVF = 1, SET = 0	IVF = 0, SET = 1
7.	IVF = 1, SET = 0	IVF = 1, SET = 0
8.	IVF = 1, SET = 0	IVF = 1, SET = 1
9.	IVF = 0, SET = 1	IVF = 0, SET = 1
10.	IVF = 0, SET = 1	IVF = 1, SET = 1
11.	IVF = 1, SET = 1	IVF = 0, SET = 1
12.	IVF = 1, SET = 1	IVF = 1, SET = 1

Certain combinations are ruled out, for instance, if Birth 1 occurs in the post SET period, the following birth (Birth 2) must also occur in the post SET period. The estimate for β_2 will capture a weighted average of the difference in salary changes of individuals who meet one of the combinations highlighted in green (those for whom $IVF \cdot SET$ "switches" between births), and the salary changes of individuals in the categories in red (those for whom $IVF \cdot SET$ remains constant across births).

The insight for our purposes is that mechanically adding mother fixed effects to the model as is often done will not generate meaningful estimates because certain combinations do not make sense if we wish to identify the impact of the SET reform. Consider birth schedule 11 which describes a mother who has one IVF birth post SET (and hence is subject to the SET reform), and follows this birth with a non-IVF birth (also post-SET). However, as birth 2 occurs without IVF, $IVF \cdot SET$ takes the value of 0 for this second birth, generating within-birth variation. Now if the SET reform has had a positive impact on salaries of women who have an IVF birth, we would expect the salary gain associated with birth one to persist following her second birth. Indeed, it may even be magnified following the second birth, given that her exposure to SET led to her first birth being less likely to be a twin birth and to her being less likely to have suffered complications. However, the mother fixed effect model considers the first birth 'exposed' to the SET reform and the second birth 'unexposed'. Thus, if we mechanically estimate that standard mother fixed effects model in equation 3, any permanent impacts on salary observed in the second birth will actually be subtracted from the impact of the reform, rather than seen as an impact of the reform.

[42]The same point holds as in the previous footnote. What we are concerned about here are the units with variation within $IVF \cdot SET_i$ and $IVF \cdot SET_j$ which can always be re-written such that the value of 1 is prior to the value of 0.

At a minimum, this group of individuals (an IVF followed by a non IVF birth, both post SET) should be removed from the sample in mother fixed effect models. However, even if this group is removed from the sample, the baseline group for which $IVF \cdot SET$ is constant across the two births is very heterogeneous, with 8 different birth profiles (indicated in red above). This includes mothers who have two IVF births post SET (group 12), and as such no variation in $IVF \cdot SET$, and who hence are mechanically (in the fixed effect model) unexposed to the reform, but in practice having been exposed twice to the reform.

A more thoughtful design is to compare treated women with matched controls defined as women with similar birth profiles, but different exposures to the reform. The following matched groups consider identical birth profiles but differential exposure to SET (numbers refer to profiles in the table above):

- 10 (variation in exposure) vs. 2 (no variation in exposure). This compares salary changes of mothers who have a first non IVF birth followed by a second IVF birth. The exposed mothers give birth to both after the SET reform, while the unexposed mothers give birth to both before the SET reform.
- 8 (variation in exposure) vs. 7 (no variation in exposure). This compares salary changes of mothers who have two IVF births. The exposed mothers gives birth to one IVF birth pre SET and one IVF birth post SET, while the unexposed mother gives birth to both IVF births pre-SET.
- 4 (variation in exposure) vs. 2 (no variation in exposure). This compares salary changes of mothers who have a first non IVF birth followed by a second IVF birth. The exposed mother gives birth to the first birth pre SET and the second birth post SET, while the unexposed mother gives birth to both in the pre-SET period.

In Table C1 we estimate the mother fixed effect model, pooling only the three exposure variation groups laid out above. This allows us to capture all fixed characteristics of mothers, and compare within-mother changes in outcomes based on exposure to SET. The results broadly support the benchmark results. Post-SET, IVF using women experience significant increases in fertility, birth weight, infant survival and earnings. We no longer see statistically significant impacts for the other outcomes. However it is hard to discern whether this is because the benchmark results for these estimates are driven by selection on unobservables, or because the mother-FE estimates are estimated on a smaller, selected sample in which there simply isn't enough variation to identify all of the benchmark model effects. The former is unlikely given the other tests we present in Section 7. The latter seems likely given that the number of observations for this test falls from about 900,000 in the benchmark to barely 5,000.

Table C1: Within-Mother Variation in Exposure to SET

	Matched Mother FE Sample			Original Estimates		
	Point Estimate	Standard Error	Observations	Point Estimate	Standard Error	Observations
Panel A: Mother Outcomes						
Mother's Earnings	16,737***	(3052)	4,016	10,872***	(2,061)	893,747
Mother's Sickness Benefits	-277	(489)	4,016	-665***	(207)	893,747
Father's Earnings	3456	(7706)	4,014	-2192	(5046)	893,793
Father's Sickness Benefits	-55.68	(371.15)	4,020	-166	(205)	893,793
Twin Birth	-0.057	(0.039)	4,004	-0.123***	0.005	895,336
Births within 9 Years	0.081*	(0.043)	4,004	0.074***	(0.010)	895,336
No Future Births	-0.073*	(0.038)	4,004	-0.072***	(0.007)	895,336
Maternal Morbidity	-0.022	(0.061)	4,020	-0.011*	(0.006)	895,336
Hospital Nights (birth)	0.828	(0.884)	3,838	-0.633***	(0.113)	874,814
Emergency C-Section	0.050	(0.046)	4,020	-0.015**	(0.006)	895,336
Panel B: Child Outcomes						
Birth weight	114.395***	40.430	5,550	194.673***	(12.043)	905,473
Gestational age	0.151	(0.152)	5,624	0.611***	(0.048)	908,232
Hospital Nights (0 years)	-0.168	(0.728)	5,528	-1.787***	(0.251)	903,605
Hospital Nights (1-10 yrs.)	0.343	(0.407)	5,624	-0.126	(0.123)	908,232
APGAR <7	-0.006	(0.012)	5,460	-0.004**	(0.002)	900,261
Severe Morbidity	-0.034	(0.024)	5,624	-0.031***	(0.005)	908,232
Infant Mortality	-0.023***	(0.008)	5,624	-0.001	(0.001)	908,232
Under 5 Mortality	0.002	(0.004)	5,624	-0.001	(0.001)	908,232

Notes: Columns 1-3 report point estimates, standard errors and observation numbers for mother fixed effect models laid out in Appendix C. This consists only of matched samples with variation in exposure to SET and IVF, and consists exclusively of mother's with two births. Mother's fixed effects capture all fixed characteristics of the mother. Mother's age fixed effects and birth date controls are also included to capture age and birth spacing respectively. Standard errors are clustered by mothers. Columns 4-6 present baseline (no mother FE) models reported in the body of the paper in Tables 1-4. * p<0.1, ** p<0.05, *** p<0.01.

D Measurement of IVF usage

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

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

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

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

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

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

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

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

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

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

This simple bias formula thus suggests that mis-classification will bias the estimate by the true diff-in-diff estimate, scaled by a parameter capturing the degree of mis-classification of the control group. In our case, given that this proportion $\frac{T_{mc}}{T_C+T_{mc}}$ is small, biases in estimates will also be small. And indeed, we can provide a back-of-the-envelope calculation of this bias using the observed values in the data. Assuming that the proportion of mis-classified IVF births is constant over time, we have that $\frac{T_{mc}}{T_C+T_{mc}} = \frac{9{,}507}{932{,}822} = 0.0102$. Now, for the case of birth weight, we can approximate the bias using values from the data as:

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

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 2 of 194.6 grams, we see that this suggests a (relatively) quite small attenuation of estimated effects.

E Heterogeneity in impacts of the SET reform

See Figure 10 for key outcomes, and see Appendix Figures A6 (for child outcomes) and A7 (for parent-level outcomes) for the rest. The headline result is that the impacts of SET are fairly pervasive by markers of the demographics, health and education of women. In this section we discuss the findings.

Consistent with SET being mandated, we see a significant decline in twin births among IVF-users across the board. Consistent with women over the age of 38 being allowed exemption, the decline is smaller in this group and this age gradient also reflects in smaller declines for women at parity 3 or higher (as these women are on average older). There are no significant differences in the decline across the other ages, or across education or BMI groups. If anything other than SET were driving the decline in twin births, it seems unlikely that it would drive a similar decline across these categories. The post-SET increase in subsequent fertility among IVF-uses is entirely driven by women at first parity. It is otherwise similar across groups with the exception of women with primary education, who display no change.[43]

The increase in birth weight and gestational age and the fall in the number of nights the child spends in hospital in the first year are across the board and of broadly similar magnitude, consistent with these indicators directly reflecting twinning. The absolute increase tends to be largest for women with tertiary education, consistent with baseline twin birth rates being higher among more educated women (Bhalotra and Clarke, 2019).[44] The overall improvement in the APGAR score is significant only for women age 25-29, with primary education and BMI at or below normal. The reduction in severe neonatal morbidity is driven by women at parity one and women age 25-34 but is pervasive across education and BMI groups. The average treatment effects were small and insignificant for child nights in hospital between the ages of 1 and 8 years, infant mortality and under-5 mortality. Breaking this down, we see no significant decline in hospital nights in any subgroup but we do see a decline in infant mortality among women with BMI at or below the normal threshold, and a decline in under-5 mortality for women at second parity and for women classified as obese.

Turning to maternal health, there is an across-the-board reduction in the number of nights women spend in hospital after birth, with the exception of the low-frequency groups of higher parity women and young (under 25) women. The largest drop is for less educated women, possibly reflecting that their baseline health is less robust. Emergency C-section rates and maternal morbid-

[43]This is not in line with the commonly observed negative association of fertility with education. Noting this sort of difference is potentially relevant for future research but, without further analysis, we can only speculate that it may reflect that women with primary education are less likely to be in stable partnerships and hence less likely to be in a position to continue fertility. They are a small fraction of the treated (IVF) group.

[44]Birth weight increases in every group while gestational age and hospital nights do not improve for women at high parity (order 3 or higher) and women under 25.

ity are correlated with hospital nights, but they exhibit more variation.[45] Sickness benefits paid to the mother decline across the board but are not statistically significant at high parity, among obese women, or among women at the two ends of the age distribution. Changes in sickness benefits paid to the father are also examined by mother characteristics (following consistent heterogeneity classifications across all outcomes). There was no average impact but, breaking this down, we see a decrease for partners of overweight women and women with secondary education, and an increase for the very small fraction of IVF-women at high parity.

The increase in women's wage earnings is pervasive across education and BMI group and evident at birth orders 1 and 2 (which cover the majority of IVF users) but only statistically significant for women age 35-38 (who are 30% of all IVF women), while being positive, but with CIs overlapping 0 at other ages. Wage earnings for fathers show no significant change in any subgroup, consistent with the averaged results discussed earlier.

[45]In considering this variation we note that heterogeneity will imply that the maternal morbidity index is a noisy sum over several component indicators, and that these three indicators of maternal health are all potentially influenced by hospital capacity and hospital policy. For all of these reasons we do not expect impacts as uniform as for indicators like fertility and birth weight. These are the patterns we see: The decline in C-section rates is most notable among women at first parity, age 25-34, low-BMI and with primary education. The decline in maternal morbidity is driven by women at first parity, age 30-34, with tertiary education– Table 1a shows that these are the modal IVF women in the case of both parity and educational level, so the decline impacts the majority of IVF users.

F Social costs and benefits of the SET reform

In this section we provide a crude estimate of the social costs and benefits of the SET reform to gain a broad sense of the relative magnitudes. This is no more than a simple back-of-the-envelope-analysis. We focus on hospital and labour market costs and benefits. We assess costs of SET vs DET. We then assess benefits in terms of cost savings to the health care budget stemming from reduced hospital nights after birth and fewer emergency C-sections, and labor market gains arising from the improved health of IVF using mothers and their children. We ignore the utility gain from improved health among both mothers and children and we ignore any general equilibrium effects associated with the labour market gains.

1. Costs to the health care sector.

a. There is a potential for an increase in the number of IVF treatments a woman undertakes after SET, since there is a higher probability that the treatment is successful in one cycle when using a double embryo transfer (DET). While the studies we quote in Section 2 find no significant difference in success rates of SET vs DET, this allows for an additional cycle with SET that involves the implantation of an additional frozen embryo. Large RCTs suggest the likelihood of live birth following SET is 27% compared to 42% in DET, but this value rises to 38% if a second transfer is then conducted (Thurin et al., 2004).[46] The public purse funds upto 3 cycles of IVF. A rough estimate of the increase in embryo transfers due to second round SETs is 42 minus 27% which gives 15%.[47] The internal cost of a SET IVF treatment is according to the Swedish health care authorities 30,100 SEK (about 3,209 US$). Since the number of IVF transfers completed following the reform was 71,687, around 70% of which were SET, this suggests a rough differential in future embryo transfers of 71,687*0.7*0.15=7,527. Based on the reference price for a SET procedure, this suggests that a cost increase of 30,100*7,527 = 226 million SEK (24 million US$).

Since the Swedish health care sector is financed by taxes, we also have to consider the social costs of raising the taxes to finance these additional costs – the Marginal Cost of Public Funds (MCF). According to Kleven and Kreiner (2003), these costs are 1.73 for Sweden. This means that the cost increase on 226 million SEK should be multiplied by 1.73, i.e., 391 million SEK (41.5 million US$).

b. Our results suggest several health improvements for mothers and children that may result in cost savings to the health care sector as a result of the SET reform. We find that the SET reform led to IVF mothers spending 0.633 fewer nights in hospital after birth. Since the cost of a hospital night in Sweden is on average 5,200 SEK (554 US$) this implies a cost reduction on 3,292 SEK per IVF mother.

[46]The success rates of 38% and 42% are not statistically significantly different.

[47]This assumes that individuals who initiated a DET cycle but did not conceive are also likely to undergo additional IVF cycles, such that the marginal change owes to the differential efficiency of SET versus SET in the first fresh embryo transfer.

Our results also suggest a 1.5 pp reduction in the probability of emergency C-sections. Since the cost difference between a vaginal delivery and an emergence C-section is 33,874 SEK (3,611 US$) there is a 508 SEK cost reduction per IVF delivery as a result of the reform.

The number of IVF deliveries after the reform for the period under study was 11,739, there is thus an overall cost reduction of 44.6 million SEK. Again considering the Marginal costs of public funds, this means that the true social cost of the additional care is 77.2 million SEK (8.2 million US$).

c. We find significant effects of the SET reform on child health, including measures of birth weight, hospital nights and severe neonatal morbidity. The only measure that is easily transferable to costs for the health care sector is hospital nights. Recognizing that we under-estimate the short term cost-savings arising from child health benefits, we compute the savings associated with fewer hospital nights. The point estimate is an average reduction of 1.787 per IVF birth implying a cost of 9,292 SEK (1.787*5,200 SEK, or 991 US$) per born child, or an aggregate cost on 123.3 million SEK (9,292*13,265 SEK, or 13.1 million US$). This implies a social cost on 213,3 million SEK (22.7 million US$).

2. Mother's labor supply after birth.

We estimate a significant increase in labor earnings of IVF mothers giving birth after the SET reform. The point estimate suggests an average annual increase of 10,872 SEK (1,159 US$) for the nine years following giving birth. This estimate suggests an increase of 86,976 SEK (9,272 US$) per woman having an IVF birth during the period under study, or an aggregate effect on 1,154 million SEK (123 million US$). The event study plot shows that the SET-led earnings increase among IVF-women is persistent.

The social costs of SET include the cost of raising tax revenues to finance income support programs for women undertaking IVF. Again, we use the marginal cost of public funds obtained from Kleven and Kreiner (2003) of 1.787. Our estimates indicate that IVF women claim lower sickness insurance benefits by an average of 665 SEK per year. We multiply this through the 9 years for which we observe women after birth. The total savings per women amount to 5,320 SEK, implying an aggregate saving of 62,6 million SEK 6,7 million US$). To get the additional social cost, and avoiding double counting, this number is multiplied by 0.73, to give 45.7 million SEK (4,9 million US$).

3. Long-term labor market effects from improved child health.

There is a large literature linking adverse health at birth to future labor market outcomes of the children. Using data on twins, Black et al. (2007) estimate a causal effect of birth weight on future earnings of the child. Their point estimate suggests that a 10 percent increase in birth weight increases average earnings by 1 percent. Our estimate is that SET results in a 6.1 percent increase in birth weight. Since the average annual earnings in Sweden are 433,200 SEK, this suggest an annual increase in earnings of 2,642 SEK (282 US$). Assuming a 40 year career and, as an approximation, assuming that wage growth and the discount rate cancel one another out exactly, we estimate an effect on individual life time earnings amounting to 105,701 SEK and an aggregate effect of 1,402 million SEK (149 million US$).

Taken together, according to our estimates, the SET reform implied a short-term fiscal cost on 58.1 million SEK (6.6 million US$) for the health care sector. Considering the Marginal cost of public funds (MCF), this corresponds to a cost to the society on 100.5 million SEK (11.7 US$). However, our calculation shows a medium and long-term 25 times as large *surplus* on 2,501 million SEK (290.8 US$), when also considering increases in women's earnings and averted benefit payments in the medium run and from child labor market returns in the long run.