Fertility treatments and the use of twin births as an instrument for fertility

Braakmann, Nils and Wildman, John

Newcastle University

4 March 2014

Fertility treatments and the use of twin births as an instrument for fertility

Nils Braakmann and John Wildman

Newcastle University*

[This version: March 3, 2014]

Abstract

Twin births are often used as an instrument for fertility in models investigating the impact of family size on labour market and child outcomes. However, a large share of twin births (24% in our sample) are the result of fertility treatments, potentially causing twin births to be endogenous and biasing estimates. Using data from the British Millennium Cohort Study we show that (a) mothers with and without fertility treatment are different, (b) twin births are still random after conditioning on fertility treatments, (c) both labour supply regressions and quantity-quality-tradeoff regressions for children's outcomes relying on the twin birth instrument appear to be biased and (d) the bias makes it less likely to find any results.

Keywords: Twin birth instrument, quantity-quality-tradeoff, labour supply, fertility

JEL-classification: C26, J13, J22,

---

* Both Newcastle University, Business School – Economics, 5 Barrack Road, Newcastle upon Tyne, NE1 4SE, UK. Email: nils.braakmann@newcastle.ac.uk; john.wildman@newcastle.ac.uk.

All estimates used Stata 13.1. Do files are available from the first author on request.

1 Introduction

Family size plays an important role for a number of economic questions. Examples include the effects of fertility on female labour supply or whether bigger families result in worse outcomes for children, i.e., the quantity-quality trade-off as predicted by models by Becker and Lewis (1973). A major empirical issue when attempting to investigate these questions is that family size is endogenous: In the case of labour supply models, both labour supply and measures of family size will be influenced by (typically unobserved) preferences for family life versus work, and there is also the possibility of reverse causality as employment outcomes partially determine the opportunity costs of children. In the case of the Becker and Lewis-quantity and quality model, family size and children's outcomes, are jointly determined in the same parental optimization process. As a consequence, researchers have resorted to finding cases where there is some exogenous variation in family size. A prominent example is the use of multiple births, i.e., the occurrence of twins or triplets, as used for example by Bronars and Grogger (1994), Jacobsen, Pierce III and Rosenbloom (1999) and Angrist and Evans (1998) in the labour supply literature and Rosenzweig and Wolpin (1980), Black, Devereux and Salvanes (2005) and Angrist, Lavy and Schlosser (2010).

In this paper we evaluate a threat to the future, though not necessarily past, validity of this instrument, namely the increasing use of fertility treatments, such as in vitro fertilization (IVF) or drug treatment with Clomiphene citrate. It is a well-established fact in the medical literature (e.g., Callahan et al., 1994; Gleicher et al., 2000; Fauser, Devroey and Macklon, 2005) that fertility treatments greatly increase the risk of multiple births. In fact, in the dataset used in this paper, we find that the probability of having either twins or triplets[1] increases from around 1% for women without fertility treatment to about 13% for women with fertility treatment. Even more worryingly, 24% of all the multiple births we observe in our sample are to women who have received fertility treatment, despite them forming only 2.6% of our sample.[2] Within the UK the use of fertility treatments has increased in most years since 1991. For IVF in 1991 there were around 8,000 cycles, by 2011 this had increased to just over 60,000 (Human Fertilisation and Embryology Authority, 2013). In 1992, in the UK, 0.3% of all babies born resulted from IVF treatment, by 2010 this had increased to 2% (Human Fertilisation and Embryology Authority, 2013).

These facts are a concern as they imply that, while multiple births are probably still more or less random conditional on having received fertility treatment, they are unlikely to be unconditionally random. Even worse, deciding to undergo fertility treatment is a choice that is likely to be correlated with a number of characteristics that also influence labour supply or investments into children – most prominently a very strong wish for children, but, as we demonstrate later in this paper, also with factors such as age, education, having worked before pregnancy, being white, marriage, family planning, complications during the pregnancy (i.e., health) and the birth weight of the first-born/only child. Given that we do not observe fertility treatments in most datasets commonly used by economists, these differences will

[1] We will generally talk about multiple births. The vast majority of these (96%) in our sample are twins with the remaining ones being triplets.

[2] In principle, a similar risk could arise for instruments based on miscarriage as in Buckles and Munnich (2011) if either miscarriages induce people to seek fertility treatment or fertility treated women are more likely to miscarry. However, we do not have data that would allow us to look into this issue.

introduce correlation between multiple births and (unobserved) determinants of fertility, which will render the instrument endogenous. It is important to be clear that in our view these issues do not invalidate some of the earlier results in the literature and might in fact not invalidate the future use of this instrument in countries or time period where fertility treatments are relatively uncommon. However, they will pose a threat for the future use of this instrument in countries where fertility treatments occur regularly and, more importantly, where multiple births resulting from fertility treatments are quantitatively important.[3]

In this paper, we use data from the British Millennium Cohort Study that follows a random sample of babies and their mothers that were born during late 2000 and 2001. We rely on the first three sweeps of this dataset. The first set of interviews (sweep I) were conducted 9 months after the respective birth, with the second (sweep II) and third (sweep III) following 3 and 5 years after birth. The dataset only contains mothers with at least on child, which is the group where the instrument has predictive power and which is also close to the sample restriction used by, e.g., Angrist and Evans (1998).[4] Section 2 describes this dataset in greater detail.

We proceed by explaining the use of multiple births as an instrument for fertility in greater detail in Section 3, where we also document pre-pregnancy

[3] Attempts to reduce the occurrence of multiple births with fertility treatments that are under way in a number of countries including the UK might also help in the future.

[4] The multiple birth instrument has no predictive power for the question whether someone has one vs. no child, as everyone who gives birth to twins or triplets will have decided to have at least one child. It has predictive power for the number of children beyond one as someone who planned to have one child will end up with two or three instead.

differences between women who received fertility treatments and those who did not. We also present comparisons of pre-pregnancy characteristics for mothers with single and multiple births with and without fertility treatment that suggest that the birth of twins or triplets is still a random event conditional on having received fertility treatment.

Subsequently, in section 4, we investigate the consequences of fertility treatments for labour supply regressions using the twin instrument. Comparisons of labour supply and other characteristics in all sweeps suggest mothers with and without fertility treatment are different, regardless of the number of children resulting from the pregnancy. We then compare first stages and labour supply regressions, i.e., second stages for five models: Our base specification is one that could be estimated using most household datasets where information on fertility treatments is missing, i.e., we just use the birth of twins or triplets as an instrument for family size on a range of outcomes related to labour supply. In a second model, we additionally condition on having received fertility treatment. As mentioned before, it seems likely that giving birth to twins or triplets is still random conditional on having received fertility treatment. A comparison of these two models thus allows us to quantify the bias in the estimates in the base model. As fertility treatments are typically unobserved in most datasets, we estimate a third model that instead conditions on a set of typically observed variables that we know to differ between women with and without fertility treatments. Results from this model allow us to make statements about whether this conditioning strategy might be a feasible approach when information on fertility treatments is lacking. Finally, given that women with and without fertility treatment are different, it is also possible that their LATEs will differ. We investigate this question by estimating separate regressions for these two groups.

Our findings suggest that the results from the first three models are qualitatively identical, i.e., the estimates always have the same sign, with only small changes in magnitude. However, the results also suggest that family size has a stronger negative effect on women who underwent fertility treatment. We also document that the effects of family size differ markedly over the three sweeps, which appears to be related to differences in child care usage (sweep II) and whether the child has already entered school (sweep III).

In section 5, we look into the consequences for regressions of measures of child quality on family size. Most of the literature has used either adult outcomes for children, such as wages, employment or completed education, or outcomes such as completed years of education at a certain age. Given that our children are only around 5 years of age, this option is not feasible. However, sweep 3 of the MCS contains data on several measures of child development routinely collected in England at the end of a child's foundation stage (ranging from the age of 3 to the age of 5).[5] Our investigation effectively proceeds in the same way as in section 4. We again compare characteristics and estimate first and second stage for children born to women with and without fertility treatment. Our findings are again similar to those of the preceding section: There are differences in the outcomes for single- and twin/triplet-born children between those whose mothers underwent fertility treatment and those who did not. First stages are somewhat similar, even though point estimates differ to some extent. The second stages paint a less rosy picture, however, as results

[5] The foundation stage is divided into stage 1 and stage 2. Stage 1 is non-compulsory, and takes place between 3 and 4 at pre-school or nursery. Foundation stage 2 takes place at school between ages of 4 and 5 and is also known as Key Stage 0. The data we use here comes from stage 2.

sometimes change considerably in magnitude. Fortunately, the results suggest that any bias arising because of the omission of controls for fertility treatment appears to make it less likely to find significant results. Conditioning on pre-pregnancy characteristics appears to lead to bias in a different direction than not conditioning at all. There are again marked differences between the children of women with and without fertility treatments in terms of the existence and magnitude of a family size penalty.

2 Data

We use data from three waves of the Millennium Cohort Study (MCS), which tracks a random sample of children (and their families) born during late 2000 and 2001 in the UK. Interviews were conducted at wave one when the children were around 9 months old, subsequent waves were conducted when the children were 3 and 6. Details on the design and sampling in the MCS can be found in Dex and Joshi (2005) and Hansen and Joshi (2007). The dataset is one of the few that we are aware of that covers fertility treatments alongside information on the mother and the development of the child.

We form two samples, one of mothers that we use to investigate labour supply and one of children that we use to look into child outcomes at age 5. For the labour supply sample we make the following sample restrictions: First, we use only cases where the mother conducted the parent interview leading to the loss of only 28 observations where the father was interviewed. Second, the MCS tracks the children born during the sampling week, not necessarily the parents, i.e., the main respondent can change in each sweep, either because the partner was interviewed or because the main carer for the child changed, for example because of adoption or death. For sweeps 2 and 3 we only use cases where the same person as in sweep 1 was interviewed, resulting in the loss of 881 (from 15,590) observations in sweep 2, 226 (from 12,984) observations in sweep 3. We also lose some observations in each sweep due to missing values (around 150 observations each in sweeps 1 and 2 and around 100 in sweep 3). Following our restrictions we have 18340 observations for sweep 1, 14460 for sweep 2 and 12581 for sweep 3.

Our main outcomes of interest are various dummies for employment status, mainly whether the mother is working, self-employed, a student or at home to care for the family, the mother's weekly working hours, calculated in two ways, either with zeros or with missing values for people not working, and finally whether she has a partner who is working. In sweep 1, we additionally have information on whether she is currently on maternity leave. Sweep 2 contains additional information on whether she uses relative/friends for childcare or whether the child attends a crèche or is cared for by a paid minder. Finally, for sweep 3 we have information on whether the child has already started school – which is the case for almost everyone (99%).

For the child outcome sample, we rely primarily on information from the Foundation Stage Profile, which is collected for all children in English schools at the end of their foundation stage – essentially an early education stage ranging from the ages of 3 to 5 – (see Qualifications and Curriculum Authority, 2003, for detailed information). It consists of a range of scales that assess competency and development in 13 areas, such as physical development, emotional development, reading, writing, calculating, the use of language or understanding of the world. Children are rated by their teachers during the foundation stage at school. The resulting scales range from 0 to 9, where higher values imply a "better" development. In principle, these are ordinal ratings, however, they are often treated as metric, e.g., when converting them into standardized scores. We use the unstandardized scores as outcomes. Note that given the children are only 5 years of age, we cannot use some of the more commonly used outcomes in the literature, such as completed education or labour market outcomes.

We merge this information with sweep 1 information on the mothers, in particular whether it was a multiple birth and whether the mother underwent fertility treatment prior to birth, and to sweep 3 information on family size. This is possible for 12,305 children from 7,010 mothers, of which 297 were either twins or triplets and of which 389 have a mother who underwent fertility treatment. Note that we lose all children in Wales, Scotland and Northern Ireland, where the foundation stage data are not collected.

In both samples, we have the same variables of interest: The first is whether the mother gave birth to twins or triplets. Almost all multiple births in the dataset are twins with only 10 cases of triplets. The latter are split equally between women with and without fertility treatment. Our sample contains 254 multiple births (i.e., twins or triplets) in sweep 1, of these 193 are in sweep 2 and 170 are in sweep 3. Our second key variable is whether the pregnancy was preceded by fertility treatment. In sweep 1, we have 478 women with fertility treatments, of these 394 remain in sweep 2 and 348 in sweep 3. The most common fertility treatment in the data is drug therapy with Clomiphene citrate, followed by various forms of in vitro fertilization. The treatment we look at in all second stage regressions is the number of children each woman has at each sweep. Note that women can have other children than the one tracked by the MCS.

Tables 1 and 2 present descriptive information on both estimation samples.

[TABLES 1 AND 2 ABOUT HERE.]

3 Twin births as an instrument for fertility

To illustrate the basic identification problems we use a causal diagram (or directed acyclic graph (DAG)) (Pearl, 2000; see Morgan and Winship, 2007, for a textbook treatment). In Figure 1 each directed edge (i.e., single headed arrow) such as the one from *family size* to *Y* represents a cause-effect-relationship between variables in the model, in the sense that the variable at the origin of the edge (start of the arrow) causes the variable at the terminus. A bidirected edge, such as the one between X_1 and X_2, represents common causes of the two factors that are not part of the model.

(FIGURE 1 ABOUT HERE.)

In figure 1 we are interested in the link between *family size* and *Y* or written as a linear equation

$$Y_i = \alpha + \tau * \text{Family size}_i + \varepsilon_{i,}. \quad (1)$$

where τ is the parameter of interest. In female labour supply regressions Y_i could either be a dummy for labour force status or some other measure of labour supply such as desired or actual working hours, while *family size*$_i$ would typically be the number of children the mother gave birth to or the number of children that live in the same household as her. In regressions concerned with the quantity-quality-trade-off for children, Y_i would be some outcome for the respective child, such as completed education or some measure of labour market outcomes, and *family size*$_i$ would be the number of children the respective parents gave birth to.

A direct estimation of this link is hindered by the presence of (potentially unobserved) sets of confounding variables, X_1 and X_2.[6] In equation (1) these would be part of ε and would render *family size*$_i$ endogenous. For example, in female labour

[6] If both X_1 and X_2 were observed, it would be possible to condition on them and use OLS, matching or other selection-on-observables estimators to look at the link between family size and the outcome.

supply models, both family size and the propensity to work will be influenced by (typically unobserved) preferences for work and family size. Furthermore, a woman's work opportunities will to some extent determine the opportunity costs of childrearing. In models of quantity-quality-trade-offs, parental investments into children (i.e., quality) and the number of children are the result of the same optimization process and are consequently jointly determined.

If, initially, we ignore the issues caused by fertility treatments, one way to proceed is to use *multiple births* as an instrument for f*amily size*. This appears to be an attractive strategy because the biological process governing whether a pregnancy results in a singleton or multiple births is outside of the control of the respective parents and thus uncorrelated with any unobserved preferences for family life, any parental optimization process, or the opportunity costs of childrearing.[7]

In figure 1, this situation is depicted in panel (a). In this scenario, *multiple births* lead to quasi-random variation in *family size* that are unrelated to the confounders X_1 and X_2 (or equivalently to ε). In this case the probability limit of the IV estimate of τ can be written as:

$$\hat{\tau} = \tau + \frac{Cov(multiple\ birth, \varepsilon)}{Cov(multiple\ birth, family\ size)} \qquad (2)$$

Equation (2) makes it clear that if multiple births and the unobservables, ε_i, from (1) are uncorrelated, the IV estimate will be consistent as *Cov(multiple births, ε)* would be zero and the bias term in equation (2) would disappear. A central condition

[7] There has been some debate about the quality of this instrument (e.g., Black, Devereux and Salvanes, 2005) as it is known that multiple births become more likely for older mothers. However, it is usually comparatively easy to account for this by conditioning on age in a flexible way, for example through age dummies.

for this to be plausible is that twin births are (more or less) random. However, with fertility treatments this is unlikely to be the case: Fertility treatments are known to cause multiple births and fertility treatments are likely to be correlated with at least some of the confounders: In many countries, fertility treatment is expensive and not fully covered by (state) health insurance, which implies that it is likely to be correlated with parental resources. These in turn matter for labour supply and parental investment into children as they determine the budget constraint and the (non-labour) income a parent can expect when not working. Furthermore, pregnancies preceded by fertility treatment are by definition always planned. They are also likely to be correlated with a strong desire for children as fertility treatments are generally preceded by a number of attempts to conceive naturally, i.e., they are generally not the first thing someone tries when trying to become pregnant.

Panel (b) of figure 1 illustrates the resulting problem: Fertility treatments create an association between *multiple births* and the confounders in X_1, i.e., multiple births are not randomly assigned. This in turn opens a backdoor path $Y \leftarrow X_1 \rightarrow fertility\ treatments \rightarrow multiple\ births\ \rightarrow family\ size\ \rightarrow Y$ between *multiple births* and the outcome. In more standard econometric terms, we can consider fertility treatments as an omitted variable. This means that the error term for equation (1) can be re-written as:

$$\varepsilon_i = \delta_1 * \text{fertility treatment}_i + \nu_i \qquad (3)$$

where δ_1 is the marginal effect of fertility treatment on labour market decisions and ν_i is a new error term that is still correlated with family size, i.e., it is likely that family size will still be endogenous after conditioning on having received fertility treatment. From (3) we can see that the covariance between multiple birth and ε_i is:

$$\text{Cov(multiple birth, } \varepsilon) = \delta_1 * \text{Cov(multiple birth, fertility treatment)} \qquad (4)$$

Using (4) we can write the plim of τ as:

$$\hat{\tau} = \tau + \delta_1 \frac{Cov(multiple\ birth, fertility\ treatment)}{Cov(multiple\ birth, family\ size)} \qquad (5)$$

Equation (5) demonstrates that bias of the IV estimate will depend on two elements: Firstly, the strength of the relationship between fertility treatments and the respective outcome (δ_1), i.e., how strongly the differences between mothers with and without fertility treatment affect the outcome of interest. And secondly, the importance of fertility treatments for the occurrence of multiple births, i.e., the covariance between multiple births and fertility treatments. This covariance is likely to be positive as the use of fertility treatments is consistently linked to multiple births in the medical literature (e.g., Callahan et al., 1994; Gleicher et al., 2000; Fauser, Devroey and Macklon, 2005). In our sample the likelihood of having multiple births is 1% for women without fertility treatment and 13% for women who had fertility treatment and 24% of all multiple births observed in the data are preceded by fertility treatments.

As an increasing number of women use fertility treatments, the second part of the bias term in (5) will become stronger as *Cov(multiple birth, fertility treatment)* will increase. It is also possible that δ_1 will change as the composition of the group of women who undergo fertility treatment changes. Furthermore, it is not possible, a priori, to sign δ_1. For example, in labour supply regressions it could be positive because fertility treatments are used by individuals with a higher propensity to work, or it could be negative as the use of fertility treatments will be correlated with a desire for children and that may be correlated with fewer individuals choosing employment. In quality-quantity-trade-off regressions, both more favourable characteristics of the mother and a stronger desire for children would suggest δ_1 to be positive. However, as fertility treatments are expensive, the budget constraints for parents with and without them might be different, which could well result in a negative δ_1.

Faced with these problems there are two ways to block the backdoor path $Y \leftarrow X_1 \rightarrow fertility\ treatments \rightarrow multiple\ births\ \rightarrow family\ size\ \rightarrow Y$ opened by the relationship between X_1, *fertility treatments* and *multiple births*. Firstly, if we observe fertility treatment, as we do, then it is possible to condition on it directly. This closes the backdoor path and removes any association between the confounders in X_1 and *multiple births*. Secondly, if all elements in X_1 were observed, one could condition on those directly, which would have an equivalent effect. A problem with this second strategy is that it is unlikely that all elements of X_1 are observed in any given dataset. However, as the first option is only available when the use of fertility treatments is observed, conditioning on variables that may be part of X_1 may be the only option when using datasets lacking this information. This strategy has its own risk as it may introduce further bias, rather than ameliorating the bias present: Theoretically, it is only clear that conditioning on the full set of confounders in X_1 would cause δ_1 to be zero and eliminate the bias. Conditioning on a subset of confounders can attenuate the problem if δ_1 shrinks towards zero as a result. However, it could also aggravate the problem: Consider a case where X_1 consists of only two variables, A and B, whose effects cancel each other out, so that δ_1 would be zero without conditioning. Conditioning on either one of them in this case would cause δ_1 to be non-zero and would actually increase bias.

In the following we estimate and compare five models. The first model uses information that would be available in most datasets and ignores the availability of information on fertility treatments, i.e., we just instrument for family size using a dummy for whether the woman gave birth to twins or triplets. The second includes a control for whether she also received fertility treatment. Estimates from this model are consistent as conditioning on fertility treatments is sufficient for the multiple births instrument to be valid. A comparison of these two models provides a picture of the size of the bias caused by unobserved fertility treatments. As a third model we condition on a set of variables that should be available in most datasets lacking information on fertility treatments, variables that could plausibly be part of X_1. These include the education of the mother, whether she worked before the pregnancy, age at birth, ethnicity and marital status.[8] A comparison of this model with the two previous models allows us to judge whether this conditioning strategy helps to attenuate any eventual bias. Finally, as women with and without fertility treatments are clearly different, we also evaluate whether the first and second stages for them are different. To do this we estimate separate models for the two groups and compare the results. All of these estimates include dummy variables for the current age of the mother in years to control for the earlier discussed age differences between single and multiple birth mothers that are apparent among women without fertility treatment in table 4.

[TABLE 3 ABOUT HERE.]

Table 3 compares the pre-pregnancy characteristics of women based on sweep 1 of the MCS. There are a range of statistically significant and economically large differences in the table that give reason for concern: Women with fertility treatment are more likely to have a (higher or first) degree, are less likely to have no qualification, are on average 4 years older at birth, are 20 percentage points more

[8] Given the relative richness of information in the MCS we could condition on additional variables. However, we deliberately restrict our choice to variables that are realistically available to researchers trying to use the multiple birth instrument with standard household data.

likely to have worked before the pregnancy or to be married, are a lot less likely to be single, are 6 percentage point less likely to be non-white, have somewhat smaller families at sweep 1 (despite the higher likelihood of multiple births), are 13 percentage points more likely to have experienced complications during pregnancy and are 47 percentage points less likely to have an unplanned pregnancy. For most of these factors it is easy to imagine a link with either labour supply or parental investments into their children.

[TABLE 4 ABOUT HERE.]

As stated before it should be possible to use multiple births as an instrument after conditioning on fertility treatments, as multiple births are probably still conditionally random. Table 4 provides some evidence on this conjecture. We compare the same characteristics as in table 3 between women with singleton and multiple births conditional on having received fertility treatment. The picture painted in this table is a lot rosier than the one in table 3: While there are still some significant differences between women with single and multiple births in each group, these are generally a lot smaller and often not statistically significant. These suggest that using multiple births as an instrument for family size might be possible as long as we are able to condition on having undergone fertility treatment.

4 Female labour supply

We begin by documenting differences in the outcomes between women with and without fertility treatments conditional on having had a single or a multiple births. Table 5 documents these differences: In general, single-birth women with and without fertility treatment appear to be quite different. Women with fertility treatment are more likely to have a working partner in all sweeps and are also significantly more likely to be working in both sweeps 1 and 2. They are also more likely to use paid childcare. The differences in employment appear to disappear by sweep 3 when most children attend school. For those who work, working hours do not appear to be too different. Women with multiple births in the two groups appear to be much more similar. While there are still differences in the probability of having a working partner in all sweeps, the gap in employment probabilities is much smaller than among single-birth women and only significantly different from zero in sweep 1. These results suggest that there are some differences between the groups that are not related to variations in family size caused by multiple births. We now evaluate whether these also lead to differences in the first and second stages of standard labour supply regressions.

[TABLE 6 ABOUT HERE.]

Table 6 begins with the first stage regressions. Consider first the two models in columns (i) and (ii). The inclusion of a control for fertility treatment clearly strengthens the relationship between multiple births and family size: The coefficient on multiple births increases by around 20% in sweeps 1 and 2 and by about 25% in sweep 3. At the same time, the first stage F-value increases substantially. Conditioning on pre-pregnancy characteristics in column (iii) strengthens the first-stage relationship, but does very little to the first-stage coefficient on multiple births.

Comparing the first stages for women with and without fertility treatment in columns (iv) and (v) reveals that the instrument is a much better predictor of family size for women with fertility treatments with much higher first stage R^2-values and equal F-values despite a much smaller sample size. Point estimates are also much larger in sweeps 2 and 3 in column (iii), further pointing towards heterogeneity between the two groups of women.

[TABLES 7,8 AND 9 ABOUT HERE.]

A bigger question is to what extent these differences matter for second stage results? Tables 7 to 9 present evidence for sweeps 1, 2 and 3 respectively. The first thing to notice is that results in columns (i) and (ii) are generally similar. Having more children lowers the propensity to be working in favour of staying at home and caring for the family. These effects also appear to be stronger when at least one of the children is young, and decline as the child ages (across sweeps 1 to 3). There also does not appear to be any effect on the working hours for those who are working. The relatively similarity of the results in these two columns suggest that the bias from omitting fertility treatments might be negligible.

The results from columns (iii) suggest that conditioning on pre-pregnancy characteristics also does not lead to substantial changes in results. However, there are several cases where the size of coefficients in column (iii) is different from those in both columns (i) and (ii). This finding highlights that conditioning on a subset of potential confounders might sometimes make matters worse.

The third thing to note from columns (iv) and (v) is that the magnitude of the effects seems to differ between women with and without fertility treatment. In general, it appears that the negative effects are much larger for women who received fertility treatment. This result is plausible as one might expect that women who underwent the trouble and (considerable) cost to undergo fertility treatment are also more likely to sacrifice part of their career to look after these children. In sum, the results suggest that despite existing behavioural differences between women with and without fertility treatment the bias in labour supply regressions relying on a multiple birth instrument appears to be comparatively small. There are however differences in the magnitude of the effects of an additional child in the two groups with the family penalty appearing to be larger for women who underwent fertility treatment.

5 Quality-quantity trade-offs

We now turn towards an investigation of potential biases in regressions of child quality at age 5 on family size. These regressions are estimated on the sample of children as described in section 2. We essentially repeat the analysis steps of the previous section, beginning with a comparison of child outcomes for women with and without fertility treatments conditional on having had a single or a multiple births. The results are presented in table 10.

[TABLE 10 ABOUT HERE.]

The evidence in table 10 suggests that children of women with fertility treatment are consistently doing better across all scales. This is also independent of whether they were part of a single birth or twins or triplets. In principle, this evidence fits the picture painted in the previous section: Women with fertility treatment are on average more likely to be employed, to have a (working) partner, to have higher education and to have smaller families, which suggests that they should also have more resources to spend on each child. Furthermore, they might be more strongly selected in terms of their desire for a child, which might also suggest that they would be willing to spend more on each child and maybe sacrifice more of their own consumption or leisure time.

[TABLE 11 ABOUT HERE.]

Table 11 again compares first stage results across our five models. The picture we obtain here is similar to the one from the previous section: The inclusion of a control for fertility treatment again strengthens the relationship between multiple births and family size as seen by changes in the coefficients and the first stage F-values. Conditioning on pre-pregnancy characteristics also strengthens the instrument, but does again very little to change the first stage coefficient relative to column (i). In fact, if anything the coefficient appears to be slightly further off the one we get in column (ii). All other results are also similar to the sweep 3 results from table 6. This is hardly surprising as these regressions are run on the same families and in fact all variables used in these regressions are taken from the mother's interview.

[TABLE 12 ABOUT HERE.]

More interesting are the second stage results in table 12. Note first that the qualitative results in columns (i) and (ii) are again similar and as expected: Larger families lead to worse child outcomes. However, comparing columns (i) and (ii) suggests that including a control for fertility treatments (sometimes substantially) increases the penalty associated with larger families. These changes are different across outcomes: While point estimates for some outcomes such as emotional development or language grow by a third to almost a half, others, such as those for shape, space and measures barely change. As a result statistical significance also sometimes changes. Conditioning instead on pre-pregnancy characteristics in columns (iii) leads to results that are even larger than the ones in column (ii), which suggests that this conditioning might have introduced a new source of bias.

Comparing columns (iv) and (v) also suggests differences between the children of women with and without fertility treatment: In general, the penalties caused by larger family size is a lot larger in absolute terms for the children of women without fertility treatment. There are also various cases where point estimates suggest positive effects for children of women with fertility treatment, while signs are always negative for the children of women without. It is not entirely clear what causes these differences (and in fact looking into these in greater detail is beyond the scope of this paper), however, one can speculate that they might be related to some of the earlier-mentioned differences between women with and without fertility treatment in terms of resources and the desire to have children.

In sum, the suggestions from these estimates are again similar to those from the preceding section: Qualitatively, the omission of fertility treatment from regressions using a multiple birth instrument does not appear to matter much. Quantitatively, the results suggest that previous results might be biased towards zero, which is reassuring both for researchers who have found significant results (which are likely to be genuine) and for those who did not (and who might now have an idea why not).

6 Conclusion

This paper evaluated the rise of fertility treatments as a threat to the commonly used multiple birth instrument for family size. Fertility treatments might threaten this identification strategy as they are linked to the occurrence of multiple births as well as to a range of characteristics that might influence labour supply or parental investments into children. Using the British Millennium Cohort Study, which allows us to distinguish between women with and without fertility treatment, we investigate the consequences of usually not being able to control for fertility treatment in both labour supply and quantity-quality-trade-off regressions for child outcomes.

We find that there are indeed differences, both in pre-pregnancy characteristics and outcomes, between women with and without fertility treatments. Conditional on having undergone fertility treatment, the birth of twins or triplets appears to be a random event. Fortunately, first stage results usually do not change much between specifications with and without controls for fertility treatments, but including fertility treatment controls appears to strengthen the first stage relationship. In labour supply regressions, the bias from omitting fertility treatment controls appears to be comparatively small in magnitude and does not affect qualitative results. For quantity-quality-trade-off regressions for child outcomes the results are also qualitatively similar, but differ sometimes considerably in magnitude. However, any bias arising from omitting fertility treatments appears to be towards zero. In all specification, conditioning instead of a set of typically observed pre-pregnancy characteristics does not appear to help very much and might in fact cause a different type of bias. In both labour supply and the quality-quantity-trade-off regressions, we find evidence that effects differ between women with and without fertility treatments (or their respective children), which might be because of higher resources among women with fertility treatments or because this group is more strongly selected in terms of the desire to have children.

References

Angrist, Joshua D. and William N. Evans (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review 88(3)*, pp. 450-477.

Angrist, Joshua, Victor Lavy and Analia Schlosser (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics 28(4)*, pp. 773-823.

Becker, Gary S. and H. Gregg Lewis (1973). On the interaction between the quantity and quality of children. *Journal of Political Economy 81(2)*, pp. S279-288.

Black, Sandra, Paul Devereux and Kjell G. Salvanes (2005). The more the merrier? The effect of family composition on children's outcomes. *The Quarterly Journal of Economics* 120(2), pp. 669-700.

Bronars, Stephen G and Jeff Grogger (1994). The economic consequences of unwed motherhood: Using twin births as a natural experiment. *American Economic Review 84(5)*, pp. 1141-56.

Buckles, Kasey S. and Elizabeth L. Munnich (2012). Birth spacing and sibling outcomes. *Journal of Human Resources 47(3)*, pp. 613-642.

Callahan, Tamara L., Janet E. Hall, Susan L. Ettner, Cindy L. Christiansen, Michel F. Greene and William F. Crowley (1994). The economic impact of multiple-gestation pregnancies and the contribution of assisted-reproduction techniques on their incidence. *New England Journal of Medicine 331(4)*, pp. 244-249.

Dex, S. and Joshi, H. (2005) *Children of the 21st century: from birth to nine months*, Policy Press, Bristol, UK

Fauser, Bard C.J.M., Paul Devroey and Nick S. Macklon (2005). Multiple birth resulting from ovarian stimulation for subfertility treatment. *The Lancet 365(9473)*, pp. 1807-1816.

Gleicher, Norbert, Denise M. Oleske, Ilan Tur-Kaspa, Andrea Vidali and Vishvanath Karande (2000). Reducing the risk of high-order multiple pregnancy after ovarian stimulation with gonadotropins. *New England Journal of Medicine 343(1)*, pp. 2-7.

Hansen, K. and Joshi, H. (2007) *Millennium Cohort Study Second Survey: a user's guide to initial findings*, Institute of Education, London, UK

Jacobsen, Joyce P., James Wishart Pearce III and Joshua L. Rosenbloom (1999). The Effects of Childbearing on Married Women's Labor Supply and Earnings:

Using Twin Births as a Natural Experiment. *Journal of Human Resources 34(3)*, pages 449-474.

Morgan, Stephen L. and Christopher Winship (2007). *Counterfactuals and causal inference*. Cambridge University Press, Cambridge.

Pearl, Judea (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge.

Rosenzweig, Mark and Kennth I. Wolpin (1980). Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica 48(1)*, pp. 227-240.

Qualifications and Curriculum Authority (2003). *Foundation Stage Profile Handbook*. London.

Table 1: Descriptive statistics labour supply sample

Variable	Observations	Mean	Std.dev.	Min.	Max.
Twin birth	18340	0.013	0.11	0	1
Triplet birth	18340	0.001	0.02	0	1
Multiple birth	18340	0.014	0.12	0	1
Had fertility treatment	18340	0.026	0.16	0	1
Pregnancy was surprising	18340	0.460	0.50	0	1
No qualification	18340	0.195	0.40	0	1
Qualification up to O-level/GCSE or equivalent	18340	0.335	0.47	0	1
A-level	18340	0.093	0.29	0	1
Higher education diploma	18340	0.084	0.28	0	1
First degree	18340	0.124	0.33	0	1
Higher degree (Master, PhD)	18340	0.033	0.18	0	1
Age at birth	18340	28.326	5.95	14	51
Had job before pregnancy	18340	0.023	0.15	0	1
Non-white ethnicity	18340	0.159	0.37	0	1
Married (1st marriage)	18340	0.555	0.50	0	1
Remarried (2nd or higher marriage)	18340	0.041	0.20	0	1
Single	18340	0.335	0.47	0	1
Divorced or separated	18340	0.068	0.25	0	1
Illness or problems during pregnancy	18340	0.378	0.48	0	1
Fertility and outcomes at time of sweep 1 interview (within 1 year of birth)					
Number of children	18340	1.953	1.09	1	10
Age	18340	29.137	5.95	14	52
Employed	18340	0.400	0.49	0	1
On maternity leave	18340	0.018	0.13	0	1
Self-employed	18340	0.026	0.16	0	1
Student	18340	0.009	0.09	0	1
At home to care for family	18340	0.542	0.50	0	1
Weekly working hours (includes 0)	18340	11.745	14.66	0	86
Weekly working hours (excludes 0)	8669	24.848	11.35	1	86
Has working partner	18340	0.724	0.45	0	1
Fertility and outcomes at time of sweep 2 interview (3 years after birth)					
Number of children	14460	2.221	1.08	1	13
Age	14460	31.854	5.85	17	54
Employed	14460	0.477	0.50	0	1
Self-employed	14460	0.008	0.09	0	1
Student	14460	0.012	0.11	0	1
At home to care for family	14460	0.437	0.50	0	1
Weekly working hours (includes 0)	14460	12.447	14.38	0	114
Weekly working hours (excludes 0)	7558	23.814	11.19	1	114
Has working partner	14460	0.752	0.43	0	1
Uses childcare by conducted by relatives/friends	14460	0.282	0.45	0	1
Uses paid childcare	14460	0.126	0.33	0	1
Fertility and outcomes at time of sweep 3 interview (5 years after birth)					
Number of children	12581	2.394	1.06	1	13
Age	12581	34.124	5.81	18	58
Employed	12581	0.527	0.50	0	1
Self-employed	12581	0.011	0.11	0	1
Student	12581	0.012	0.11	0	1
At home to care for family	12581	0.371	0.48	0	1
Weekly working hours (includes 0)	12581	13.955	14.46	0	100
Weekly working hours (excludes 0)	7390	23.758	11.10	0	100
Has working partner	12581	0.751	0.43	0	1
Child attends school	12581	0.988	0.11	0	1

Table 2: Descriptive statistics, child outcome sample (England only)

Variable	Mean	Std. Dev.	Min	Max
Child development scales (all 0 to 9)				
Disposition and attitudes	7.311	1.38	0	9
Social development	6.823	1.61	0	9
Emotional development	6.844	1.76	0	9
Language for communication and thinking	6.753	1.75	0	9
Linking sounds and letters	6.147	2.14	0	9
Reading	6.401	1.81	0	9
Writing	5.824	2.05	0	9
Numbers as labels and for counting	7.242	1.51	0	9
Calculating	6.277	1.96	0	9
Shape, space and measures	6.765	1.65	0	9
Knowledge and understanding of the world	6.677	1.65	0	9
Physical development	7.214	1.43	0	9
Creative development	6.661	1.53	0	9
Twin birth	0.0238	0.15	0	1
Triplet birth	0.001	0.02	0	1
Multiple birth	0.024	0.15	0	1
Had fertility treatment	0.032	0.17	0	1
Number of children sweep 3	2.307	1.01	1	13
Observations		12,305		

Table 3: Comparison of pre-pregnancy characteristics of women with and without fertility-treatment

Variable	Without fertility treatment		With fertility treatment		P-Value means different[a]
	Mean	Std.dev.	Mean	Std.dev.	
Twin birth	0.01	0.10	0.12	0.32	0.0000
Triplet birth	0.00	0.02	0.01	0.10	0.0294
Multiple birth	0.01	0.10	0.13	0.33	0.0000
Pregnancy was surprising	0.47	0.50	0.00	0.00	0.0000
Birth weight 1st child (kg)	3.36	0.57	3.19	0.65	0.0000
Number of children at sweep 1 interview	1.96	1.09	1.54	0.75	0.0000
No qualification	0.20	0.40	0.12	0.32	0.0000
Qualification up to O-level/GCSE or equivalent	0.34	0.47	0.33	0.47	0.9835
A-level	0.09	0.29	0.11	0.31	0.3286
Higher education diploma	0.08	0.28	0.09	0.29	0.4430
First degree	0.12	0.33	0.18	0.38	0.0026
Higher degree (Master, PhD)	0.03	0.18	0.08	0.27	0.0002
Age at birth	28.22	5.94	32.29	4.94	0.0000
Had job before pregnancy	0.62	0.49	0.81	0.39	0.0000
Non-white ethnicity	0.16	0.37	0.10	0.30	0.0000
Married (1st marriage)	0.55	0.50	0.76	0.43	0.0000
Remarried (2nd or higher marriage)	0.04	0.20	0.06	0.25	0.0302
Single	0.34	0.47	0.12	0.32	0.0000
Divorced or separated	0.07	0.25	0.06	0.24	0.4015
Illness or problems during pregnancy	0.37	0.48	0.50	0.50	0.0000
Observations	17862		478		

[a] Based on two sample t-test with unequal variances.

Table 4: Comparison of pre-pregnancy characteristics of women with single and multiple births by fertility treatment

	Women without fertility treatment					Women with fertility treatment				
	Single birth		Multiple birth		P-Value means	Single birth		Multiple birth		P-value means
	Mean	Std.dev.	Mean	Std.dev.	different	Mean	Std.dev.	Mean	Std.dev.	different
Pregnancy was surprising	0.47	0.50	0.49	0.50	0.5801	0.00	0.00	0.00	0.00	
Birth weight 1st child (kg)	3.37	0.56	2.44	0.52	0.0000	3.30	0.58	2.42	0.59	0.0000
No qualification	0.20	0.40	0.23	0.42	0.2319	0.12	0.33	0.10	0.30	0.5671
Qualification up to O-level/GCSE or equivalent	0.34	0.47	0.31	0.46	0.4655	0.33	0.47	0.38	0.49	0.4690
A-level	0.09	0.29	0.06	0.24	0.0804	0.10	0.30	0.16	0.37	0.1934
Higher education diploma	0.08	0.28	0.14	0.35	0.0250	0.10	0.30	0.07	0.25	0.3539
First degree	0.12	0.33	0.13	0.34	0.6152	0.18	0.38	0.16	0.37	0.7929
Higher degree (Master, PhD)	0.03	0.18	0.03	0.16	0.5826	0.08	0.27	0.07	0.25	0.6464
Age at birth	28.20	5.94	30.12	5.70	0.0000	32.21	4.92	32.87	5.09	0.3433
Had job before pregnancy	0.62	0.49	0.64	0.48	0.5791	0.81	0.39	0.80	0.40	0.8603
Non-white ethnicity	0.16	0.37	0.12	0.32	0.0799	0.11	0.31	0.07	0.25	0.2613
Married (1st marriage)	0.55	0.50	0.59	0.49	0.2460	0.75	0.44	0.85	0.36	0.0376
Remarried (2nd or higher marriage)	0.04	0.20	0.06	0.24	0.2043	0.07	0.25	0.03	0.18	0.1630
Single	0.34	0.47	0.28	0.45	0.0871	0.12	0.33	0.05	0.22	0.0212
Divorced or separated	0.07	0.25	0.06	0.24	0.7486	0.06	0.23	0.07	0.25	0.8138
Illness or problems during pregnancy	0.37	0.48	0.46	0.50	0.0238	0.49	0.50	0.54	0.50	0.4754
Observations	17,669		193			417		61		

Table 5: Comparisons of outcomes for women with and without fertility treatment with same number of children born

	Single births					Multiple births				
	No FT		FT		P-value means different	No FT		FT		P-value means different
	Mean	Std.dev.	Mean	Std.dev.		Mean	Std.dev.	Mean	Std.dev.	
Sweep I outcomes										
Employed	0.40	0.49	0.52	0.50	0.0000	0.31	0.46	0.43	0.50	0.1127
On maternity leave	0.03	0.16	0.06	0.23	0.0078	0.02	0.14	0.03	0.18	0.6332
Self-employed	0.02	0.13	0.04	0.19	0.0222	0.04	0.19	0.08	0.28	0.2315
Student	0.01	0.09	0.00	0.07	0.2716	0.01	0.07	0.02	0.13	0.5164
At home to care for family	0.54	0.50	0.38	0.49	0.0000	0.63	0.48	0.44	0.50	0.0132
Weekly working hours (includes 0)	11.63	14.61	16.78	15.62	0.0000	9.98	14.49	15.36	15.18	0.0166
Weekly working hours (excludes 0)	24.81	11.33	25.82	11.90	0.1726	24.39	12.69	26.77	9.61	0.2744
Has working partner	0.72	0.45	0.90	0.29	0.0000	0.74	0.44	0.89	0.32	0.0047
Observations	17,669		417			193		61		
Sweep II outcomes										
Employed	0.47	0.50	0.59	0.49	0.0000	0.46	0.50	0.43	0.50	0.6834
Self-employed	0.01	0.09	0.00	0.05	0.0810	0.00	0.00	0.00	0.00	n/a
Student	0.01	0.11	0.01	0.08	0.1192	0.01	0.12	0.06	0.24	0.2026
At home to care for family	0.44	0.50	0.31	0.47	0.0000	0.46	0.50	0.41	0.50	0.5159
Weekly working hours (includes 0)	12.38	14.36	15.45	15.01	0.0002	11.73	14.35	12.41	14.23	0.7711
Weekly working hours (excludes 0)	23.83	11.14	23.34	12.49	0.5595	23.14	11.89	24.35	10.13	0.6217
Has working partner	0.75	0.43	0.92	0.28	0.0000	0.77	0.42	0.82	0.39	0.4496
Uses childcare by conducted by relatives/friends	0.28	0.45	0.29	0.46	0.7216	0.23	0.42	0.16	0.37	0.2744
Uses paid childcare	0.12	0.33	0.22	0.42	0.0000	0.08	0.27	0.18	0.39	0.0947
Observations	13,942		343			142		51		
Sweep III outcomes										
Employed	0.53	0.50	0.57	0.50	0.1298	0.54	0.50	0.52	0.50	0.8149
Self-employed	0.01	0.11	0.00	0.06	0.0195	0.01	0.09	0.00	0.00	0.3193
Student	0.01	0.11	0.01	0.08	0.2225	0.01	0.09	0.02	0.14	0.5745
At home to care for family	0.37	0.48	0.30	0.46	0.0067	0.34	0.48	0.31	0.47	0.6931
Weekly working hours (includes 0)	13.92	14.48	15.32	13.71	0.0817	13.68	14.32	15.48	14.16	0.4594
Weekly working hours (excludes 0)	23.80	11.11	22.53	10.67	0.0950	22.86	11.47	23.97	10.20	0.6284
Has working partner	0.75	0.43	0.89	0.32	0.0000	0.80	0.41	0.88	0.33	0.1902
Child attends school	0.99	0.11	0.99	0.11	0.8490	0.98	0.13	0.96	0.20	0.4232
Observations	12,111		300			122		48		

Table 6: First stage results, labour supply sample

	(i) All women	(ii) All women, control for fertility treatment	(iii) All women, control for pre-pregnancy characteristics	(iv) Only women with fertility treatment	(v) Only women without fertility treatment
			Sweep I		
Multiple birth (1 = yes)	0.882*** (0.072)	1.042*** (0.071)	0.885*** (0.063)	1.101*** (0.094)	1.028*** (0.086)
Fertility treatment (1 = yes)		-0.784*** (0.034)			
R^2	0.010	0.024	0.227	0.248	0.011
Kleinbergen-Paap F-stat	149.13	215.69	194.46	136.49	141.21
Observations	18340	18340		478	17862
			Sweep II		
Multiple birth (1 = yes)	0.685*** (0.077)	0.830*** (0.077)	0.694*** (0.067)	1.007*** (0.121)	0.787*** (0.093)
Fertility treatment (1 = yes)		-0.643*** (0.042)			
R^2	0.006	0.015	0.191	0.180	0.006
Kleinbergen-Paap F-stat	78.94	116.27	107.69	69.08	71.46
Observations	14460	14460		394	14066
			Sweep III		
Multiple birth (1 = yes)	0.573*** (0.083)	0.715*** (0.084)	0.605*** (0.073)	0.903*** (0.127)	0.665*** (0.103)
Fertility treatment (1 = yes)		-0.580*** (0.046)			
R^2	0.004	0.012	0.157	0.138	0.004
Kleinbergen-Paap F-stat	48.04	73.06	68.05	50.48	41.57
Observations	12581	12581	12581	348	12233

Coefficient, robust standard errors in parentheses. */**/*** denote statistical significance on the 10%, 5% and 1% level respectively. All estimates include age in years as dummies. Column (iii) also contains dummies for various completed qualifications, age at birth, a dummy for having worked before the pregnancy, a dummy for non-white ethnicity and dummy variables for marital status.

Table 7: Outcomes Sweep I interview (within 1 year of birth)

	(i) All women	(ii) All women, control for fertility treatment	(iii) All women, control for pre-pregnancy characteristics	(iv) Only women with fertility treatment	(v) Only women without fertility treatment
			Employed (1 = yes)		
Number of children	-0.107***	-0.106***	-0.122***	-0.096	-0.107***
	(0.033)	(0.028)	(0.031)	(0.061)	(0.031)
			Self-employed (1= yes)		
Number of children	-0.010	-0.012	-0.010	-0.018	-0.009
	(0.011)	(0.009)	(0.011)	(0.023)	(0.010)
			On maternity/parental leave (1 = yes)		
Number of children	0.025*	0.018	0.025*	0.029	0.016
	(0.015)	(0.012)	(0.014)	(0.032)	(0.013)
			Fulltime student (1 = yes)		
Number of children	0.003	0.002	0.004	0.007	-0.001
	(0.006)	(0.005)	(0.006)	(0.014)	(0.005)
			At home and caring for family (1 = yes)		
Number of children	0.094***	0.101***	0.110***	0.078	0.106***
	(0.033)	(0.029)	(0.030)	(0.060)	(0.032)
			Weekly working hours (includes 0 for those not working)		
Number of children	-1.997**	-2.340***	-2.297***	-1.917	-2.452**
	(0.998)	(0.851)	(0.890)	(1.833)	(0.958)
			Weekly working hours (excludes those not working)		
Number of children	-0.020	-0.185	0.110	0.391	-0.606
	(1.275)	(1.118)	(1.269)	(1.702)	(1.385)
			Has a working partner (1= yes)		
Number of children	-0.005	-0.025	-0.025	-0.032	-0.022
	(0.029)	(0.024)	(0.024)	(0.039)	(0.029)
Observations (all but second working hours regression)	18340	18340	18340	478	17862
Observations (second working hours regression)	8669	8669	8669	306	8363

Coefficient, robust standard errors in parentheses. */**/*** denote statistical significance on the 10%, 5% and 1% level respectively. All estimates include age in years as dummies. Column (ii) additionally contains a dummy for having received fertility-treatment. Column (iii) also contains dummies for various completed qualifications, age at birth, a dummy for having worked before the pregnancy, a dummy for non-white ethnicity and dummy variables for marital status.

Table 8: Outcomes Sweep II interview (3 years after birth)

	(i) All women	(ii) All women, control for fertility treatment	(iii) All women, control for pre-pregnancy characteristics	(iv) Only women with fertility treatment	(v) Only women without fertility treatment
	Employed (1 = yes)				
Number of children	-0.082	-0.083*	-0.100**	-0.194***	-0.044
	(0.052)	(0.044)	(0.049)	(0.073)	(0.053)
	Self-employed (1 = yes)				
Number of children	-0.009***	-0.007***	-0.008***	-0.002	-0.008***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
	Fulltime student (1 = yes)				
Number of children	0.025	0.019	0.025	0.057*	0.005
	(0.017)	(0.013)	(0.017)	(0.033)	(0.013)
	At home and caring for family (1 = yes)				
Number of children	0.073	0.081*	0.091*	0.118	0.065
	(0.052)	(0.043)	(0.048)	(0.073)	(0.053)
	Weekly working hours (includes 0 for those not working)				
Number of children	-2.468*	-2.418**	-2.698**	-4.004**	-1.827
	(1.466)	(1.228)	(1.350)	(2.037)	(1.491)
	Weekly working hours (excludes those not working)				
Number of children	-0.813	-0.551	-0.471	-1.077	-0.909
	(1.613)	(1.428)	(1.590)	(2.217)	(1.762)
	Has a working partner (1= yes)				
Number of children	-0.012	-0.035	-0.036	-0.078	-0.014
	(0.043)	(0.036)	(0.039)	(0.056)	(0.044)
Observations (all except below)	14460	14460	14460	394	14066
Observations (second working hours regression)	7558	7558	7558	253	7305

Coefficient, robust standard errors in parentheses. */**/*** denote statistical significance on the 10%, 5% and 1% level respectively. All estimates include age in years as dummies. Column (ii) additionally contains a dummy for having received fertility-treatment. Column (iii) also contains dummies for various completed qualifications, age at birth, a dummy for having worked before the pregnancy, a dummy for non-white ethnicity and dummy variables for marital status.

Table 9: Outcomes Sweep III interview (5 years after birth)

	(i) All women	(ii) All women, control for fertility treatment	(iii) All women, control for pre-pregnancy characteristics	(iv) Only women with fertility treatment	(v) Only women without fertility treatment
			Employed (1 = yes)		
Number of children	-0.042	-0.032	-0.078	-0.084	-0.013
	(0.066)	(0.054)	(0.061)	(0.086)	(0.067)
			Self-employed (1 = yes)		
Number of children	-0.005	-0.003	-0.003	-0.002	-0.002
	(0.011)	(0.009)	(0.010)	(0.002)	(0.012)
			Fulltime student (1 = yes)		
Number of children	0.002	0.002	0.003	0.026	-0.004
	(0.015)	(0.012)	(0.014)	(0.025)	(0.012)
			At home and caring for family (1 = yes)		
Number of children	0.001	0.008	0.041	0.029	-0.000
	(0.063)	(0.051)	(0.055)	(0.084)	(0.064)
			Weekly working hours (includes 0 for those not working)		
Number of children	-1.586	-1.282	-2.370	-0.545	-1.618
	(1.879)	(1.532)	(1.688)	(2.452)	(1.912)
			Weekly working hours (excludes those not working)		
Number of children	-1.272	-0.646	-1.002	0.804	-1.424
	(1.721)	(1.461)	(1.697)	(2.335)	(1.800)
			Has a working partner (1= yes)		
Number of children	0.048	0.011	0.006	-0.030	0.025
	(0.052)	(0.042)	(0.048)	(0.058)	(0.054)
Observations (all but second working hours regression)	12581	12581	12581	348	12233
Observations (second working hours regression)	7390	7390	7390	235	7155

Coefficient, robust standard errors in parentheses. */**/*** denote statistical significance on the 10%, 5% and 1% level respectively. All estimates include age in years as dummies. Column (ii) additionally contains a dummy for having received fertility-treatment. Column (iii) also contains dummies for various completed qualifications, age at birth, a dummy for having worked before the pregnancy, a dummy for non-white ethnicity and dummy variables for marital status.

Table 10: Comparisons of child outcomes for women with and without fertility treatment with same number of children born

	Single births					Multiple births				
	No FT		FT		P-value means different	No FT		FT		P-value means different
	Mean	Std.dev	Mean	Std.dev		Mean	Std.dev	Mean	Std.dev	
Disposition and attitudes	7.30	1.38	7.58	1.16	0.0000	7.16	1.69	7.63	1.47	0.0161
Social development	6.81	1.61	7.07	1.48	0.0014	6.70	1.85	7.59	1.40	0.0000
Emotional development	6.84	1.77	7.17	1.54	0.0001	6.61	1.94	7.21	1.60	0.0057
Language for communication and thinking	6.74	1.75	7.13	1.54	0.0000	6.47	2.01	7.35	1.64	0.0001
Linking sounds and letters	6.14	2.14	6.61	1.86	0.0000	5.77	2.32	6.48	2.33	0.0177
Reading	6.39	1.81	6.92	1.57	0.0000	6.08	2.00	6.60	1.72	0.0235
Writing	5.81	2.05	6.22	1.88	0.0001	5.70	2.15	5.93	2.20	0.4166
Numbers as labels and for counting	7.24	1.51	7.48	1.23	0.0003	7.00	1.95	7.52	1.58	0.0148
Calculating	6.27	1.96	6.67	1.67	0.0000	5.92	2.18	6.50	2.21	0.0402
Shape, space and measures	6.76	1.65	6.98	1.45	0.0055	6.47	1.88	6.74	1.74	0.2209
Knowledge and understanding of the world	6.67	1.64	6.99	1.44	0.0000	6.59	1.97	7.15	1.38	0.0048
Physical development	7.21	1.43	7.43	1.29	0.0019	7.05	1.78	7.24	1.64	0.3540
Creative development	6.65	1.53	7.00	1.33	0.0000	6.63	1.71	7.07	1.46	0.0236
Observations	13,239		358			257		82		

All outcomes are scales from 0 to 9. Higher values imply higher development.

Table 11: First stages, child outcome sample

	(i) All women	(ii) All women, control for fertility treatment	(iii) All women, control for pre-pregnancy characteristics	(iv) Only women with fertility treatment	(v) Only women without fertility treatment
			Sweep III		
Multiple birth (1 = yes)	0.552***	0.652***	0.512***	0.723***	0.632***
	(0.064)	(0.065)	(0.056)	(0.098)	(0.078)
Fertility treatment (1 = yes)		-0.447*** (0.041)			
R^2	0.007	0.013	0.173	0.124	0.007
Kleinbergen-Paap F-stat	75.400	100.973	83.34	54.718	65.431
Observations	12305	12305	12305	389	11916

Coefficient, robust standard errors in parentheses. */**/*** denote statistical significance on the 10%, 5% and 1% level respectively. All estimates include age in years as dummies. Column (ii) additionally contains a dummy for having received fertility-treatment. Column (iii) also contains dummies for various completed qualifications, age at birth, a dummy for having worked before the pregnancy, a dummy for non-white ethnicity and dummy variables for marital status.

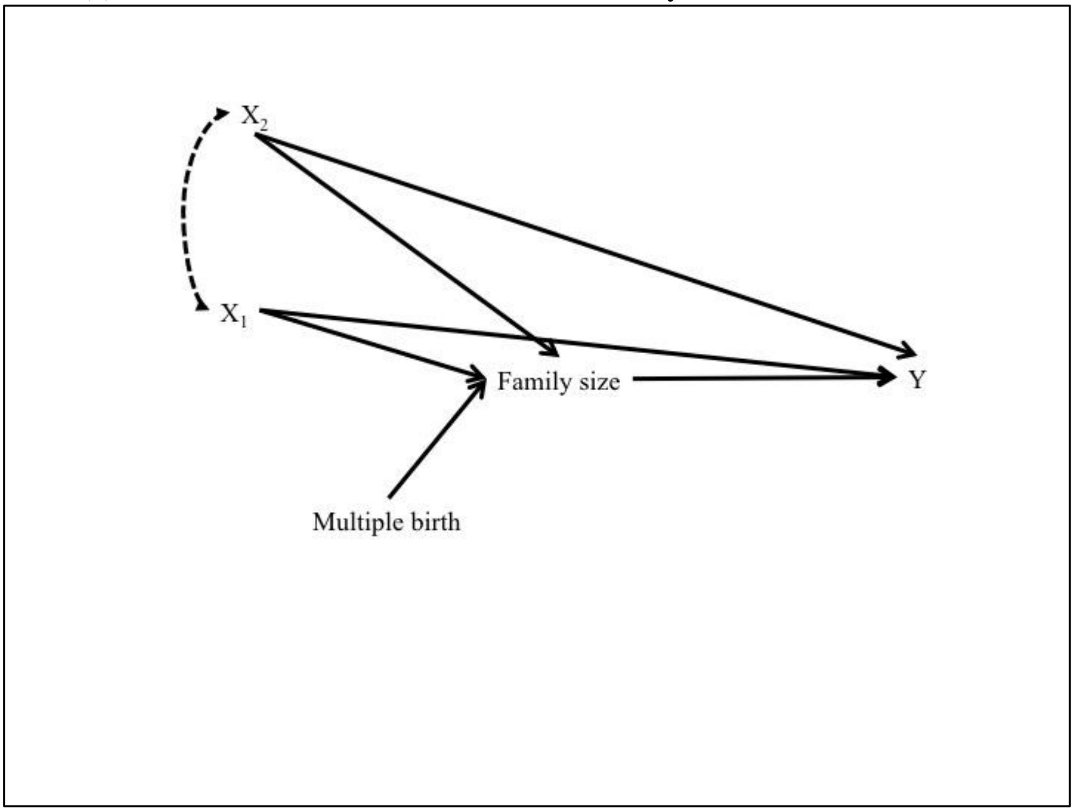
Table 12: Family size and child development outcomes, second stage results, child outcome sample

	(i) All women	(ii) All women, control for fertility treatment	(iii) All women, control for pre-pregnancy characteristics	(iv) Only women with fertility treatment	(v) Only women without fertility treatment
		Disposition and attitudes			
Number of children	-0.092	-0.210	-0.236	-0.108	-0.241
	(0.171)	(0.145)	(0.178)	(0.255)	(0.172)
		Social development			
Number of children	0.157	0.005	-0.031	0.645***	-0.195
	(0.191)	(0.160)	(0.194)	(0.240)	(0.190)
			Emotional development		
Number of children	-0.276	-0.381**	-0.514**	-0.003	-0.500**
	(0.200)	(0.174)	(0.217)	(0.261)	(0.213)
		Language for communication and thinking			
Number of children	-0.178	-0.324*	-0.396*	0.205	-0.490**
	(0.201)	(0.170)	(0.207)	(0.257)	(0.204)
		Linking sounds and letters			
Number of children	-0.484**	-0.591***	-0.762***	-0.161	-0.726***
	(0.239)	(0.203)	(0.241)	(0.387)	(0.233)
		Reading			
Number of children	-0.484**	-0.602***	-0.772***	-0.511*	-0.631***
	(0.200)	(0.172)	(0.205)	(0.309)	(0.204)
		Writing			
Number of children	-0.110	-0.251	-0.361*	-0.345	-0.221
	(0.219)	(0.185)	(0.219)	(0.365)	(0.216)
		Numbers as labels and for counting			
Number of children	-0.238	-0.317*	-0.416**	0.140	-0.460**
	(0.193)	(0.166)	(0.197)	(0.221)	(0.203)
		Calculating			
Number of children	-0.551**	-0.635***	-0.831***	-0.196	-0.772***
	(0.227)	(0.194)	(0.239)	(0.358)	(0.225)
			Shape, space and measures		
Number of children	- 0.512***	-0.531***	-0.745***	-0.291	-0.606***
	(0.189)	(0.163)	(0.204)	(0.276)	(0.195)
		Knowledge and understanding of the world			
Number of children	-0.028	-0.183	-0.222	0.020	-0.247
	(0.189)	(0.161)	(0.191)	(0.200)	(0.199)
		Physical development			
Number of children	-0.194	-0.270*	-0.337*	-0.461*	-0.210
	(0.178)	(0.152)	(0.187)	(0.279)	(0.182)
		Creative development			
Number of children	-0.029	-0.170	-0.188*	-0.014	-0.219
	(0.170)	(0.145)	(0.172)	(0.222)	(0.176)
Observations	12305	12305	12305	389	11916

Coefficient, robust standard errors in parentheses. */**/*** denote statistical significance on the 10%, 5% and 1% level respectively. All estimates include age in years as dummies. Column (ii) additionally contains a dummy for having received fertility-treatment. Column (iii) also contains dummies for various completed qualifications, age at birth, a dummy for having worked before the pregnancy, a dummy for non-white ethnicity and dummy variables for marital status.

Figure 1: Causal diagram for the multiple birth instrument with and without fertility treatments

Panel (a): The twin births instrument without fertility treatments



Panel (b): The twin births instrument with fertility treatments