Twin Birth and Maternal Condition[*]

Sonia Bhalotra[†]   Damian Clarke[‡]

July 9, 2018

Abstract

Twin births are often construed as a natural experiment in the social and natural sciences on the premise that the occurrence of twins is quasi-random. We present population-level evidence that challenges this premise. Using individual data for 17

million births in 72 countries, we demonstrate that indicators of mother's health, health-related behaviours and the prenatal environment are systematically positively associated with twin birth. The associations are sizeable, evident in richer and poorer countries, evident even among women who do not use IVF, and hold for numerous different measures of health. We discuss potential mechanisms, showing evidence that favours selective miscarriage.

1   Introduction

Twins have intrigued humankind for more than a century (Thorndike, 1905). In behavioural genetics, demography and psychology, monozygotic twins are studied to assess the importance of nurture relative to nature (Polderman et al., 2015). In the social sciences, twin births are also used to denote an unexpected increase in family size which assists causal identification of the impact of fertility on investments in children and on women's labour supply (Rosenzweig and Wolpin, 2000, 1980a; Bronars and Grogger, 1994; Black et al., 2005). A premise of studies that use twin differences or the twin instrument is that twin births are quasi-random and have no direct impact (except through fertility) on the outcome under study. We present new population-level evidence that challenges this premise. Using 16,962,165 births in 72 countries, of which 462,246 (2.73%) are twins, we show that the likelihood of a twin birth varies systematically with maternal condition.

In particular, our estimates establish that mothers of twins are selectively healthy.[1]

We document that the association of twin births and maternal condition is meaningfully large, and widespread. We show that is evident in richer and poorer countries, and that it holds for sixteen different markers of maternal condition 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.[2] The effects are sizeable, with a 1 standard deviation improvement in the indicator tending to increase the likelihood of twinning by 6-12%.

Previous research has documented that twins have different endowments from singletons, for example, twins are more likely to have low birth weight and congenital anomalies (Hall, 2003; Rosenzweig and Zhang, 2009). We focus not on differences between twins and singletons but rather on differences between mothers of twins and singletons, which indicate whether occurrence of twin births is quasi-random. It is known that twin births are not strictly random, occurring more frequently among older mothers, at higher parity and in certain races and ethnicities (Hall, 2003; Bulmer, 1970), but as these variables are typically observable, they can be adjusted for (as in Rosenzweig and Wolpin (1980a)).[3] Similarly, it is well-documented that women using artificial reproductive technologies (ART) are more likely to give birth to twins (Vitthala et al., 2009) but

---

[1]Twins are not as rare as we may think: 1 in 80 live births and hence 1 in 40 newborns is a twin. In general and, for instance, in the United States (US), there is a positive trend in twin births.

[2]We also show that a positive association of the chances of having twins with health-related behaviours in pregnancy (healthy diet, smoking, alcohol, drug consumption), although we do not rely upon this because behaviours in pregnancy may reflect a response to the mother's knowledge that she is carrying twins.

[3]Other correlates identified in the medical literature but not reflected in social science research

ART-use is recorded in many birth registries, and so it can be controlled for and a conditional randomness assumption upheld (Cáceres-Delpiano, 2006; Angrist et al., 2010). The reason that our finding is potentially a major challenge is that maternal condition is multi-dimensional and almost impossible to fully measure and adjust for. To take a few examples, foetal health is potentially a function of whether pregnant women skip breakfast (Mazumder and Seeskin, 2015), whether they suffer bereavement in pregnancy (Black et al., 2016), or exposure to air pollution (Chay and Greenstone, 2003).

Our underlying hypothesis is that twins are more demanding of maternal resources than singletons and, as a result, conditions that challenge maternal health are more likely to result in miscarriage of twins than of singletons. We discuss the role of alternative mechanisms including non-random conception and maternal survival selection. We provide evidence in favour of the selective miscarriage mechanism using US Vital Statistics data for 14 to 16 million births. Selective miscarriage is similarly the mechanism behind the stylized fact that weaker maternal condition is associated with a lower probability of male birth (Trivers and Willard, 1973; Almond and Edlund, 2007). We confirm this in our data, showing that twin births are more likely to be female. Our findings add a novel twist to a recent literature documenting that a mother's health and her environmental exposure to nutritional or other stresses during pregnancy influence birth outcomes, with many studies documenting lower birth weight (Currie and Moretti, 2007; Bernstein et al., 2005;

_____

include high concentrations of follicle-stimulating hormone in women, season and seasonal light, height, urbanization, and starvation (Hall, 2003) with mixed results (based on small samples) when considering social class (Campbell et al., 1974; Campbell, 1998). These results have not been documented in the economics or social science literature. In our discussion of Mechanisms we shall discuss the difference between monozygotic and dyzygotic twins.

Quintana-Domeque and Ródenas-Serrano, 2017). If birth weight is the intensive margin, we may think of miscarriage as an extensive margin response, or the limiting case of low birth weight.

Our findings have implications for research that has exploited the assumed randomness of twin births. Studies using twins to isolate exogenous variation in fertility will tend to under-estimate the impact of fertility on parental investments in children, and on women's labour supply if selectively healthy mothers invest more in children post-birth, and are more likely to participate in the labour market (as discussed in Bloom et al. (2015)). In Table 1 we summarize studies using twin births to instrument fertility, documenting the mother-level controls in each study. In some cases the validity of the conditional randomness assumption is directly probed, for instance, with respect to mother's education (Black et al. (2005), Li et al. (2008), Rosenzweig and Zhang (2009)). However, as is acknowledged in each case, any such tests are at best partial evidence in support of instrumental validity. Importantly, *no* previous study has attempted to control for maternal health conditions or behaviours. This is pertinent as it could resolve the ambiguity of the available evidence on the impacts of fertility. In particular, recent studies using the twin instrument challenge a long-standing theoretical prior of Becker and Lewis (1973) in rejecting the presence of a quantity–quality (QQ) fertility trade-off in developed countries (Black et al., 2005; Angrist et al., 2010), but our estimates suggest that this rejection could in principle arise from ignoring the positive selection of women into twin birth. Similarly, research using the twin instrument tends to find that additional children have relatively little influence on female labour force participation (FLFP), see Lundborg et al. (2017). But, again, these estimates are likely to be downward biased. The results of studies in Economics, Psychology, Education and Biology that instead exploit the genetic similarity of twins will not be biased but will tend to have more restricted external validity than previously assumed.[4]

---

[4]The twin instrument has been criticised for other reasons. A recent critique of the use of twins

2 Methodology

In this section we discuss two distinct approaches to testing our hypothesis that twins are selectively born to healthier mothers. We identify variation in the mother's health before she gives birth to twins, and before she knows she will give birth to twins. In the first approach we use information on her health condition (morbidities, height, weight), health-related behaviours, access to health care and environmental health stressors. In our second approach we use as a marker of maternal health the foetal or infant survival rate of her births prior to the birth at which she has twins (with parity-matched counterfactuals). The methods used to investigate potential mechanisms driving this are discussed later.

We conduct three robustness checks. First, we restrict the sample to non-ART births. It is important to demonstrate that our hypothesis holds independently of ART use because there is a positive association of ART with the likelihood of twin births (Vitthala et al., 2009), and ART users are typically more educated and wealthy (Lundborg et al., 2017). Another potential concern is that we are capturing genetic traits that, for instance, are associated with the woman's height or weight, and also correlated with her predisposition toward twin birth. This would appear to be

---

to identify the QQ trade-off has argued that parental behaviours may respond to the endowment of twins and not only to the fact that twin births represent a fertility shock. Rosenzweig and Zhang (2009) highlight that twins have lower birth endowments. They argue that if parents reinforce endowments then they may reallocate resources towards the better endowed children born before the twins, obscuring any underlying QQ trade-off; and this is examined in Angrist et al. (2010) and Fitzsimons and Malde (2014). We remain agnostic on this. Our critique is in principle orthogonal to this critique, providing a different reason that an underlying QQ trade-off may be obscured, relating to endowments and behaviours of mothers. This critique has not been previously considered.

a second-order concern since we do not only rely upon woman-specific measures of health but also show a positive association of twinning with environmental stressors, health-facilities and health-related behaviours. We nevertheless investigate this concern in two different ways. First we test whether we can identify a positive association of the probability that a birth is a twin with woman-specific time-varying health indicators conditional upon woman fixed effects that sweep out genetic influences. Second, we leverage biomedical research showing that monozygotic (MZ) twins are randomly allocated across mothers, although genetic predispositions may influence the chances of having dizygotic (DZ) twins (Meulemans et al., 1996). Ideally, we would restrict the sample to MZ twins, but MZ vs DZ are not identified in the data. Instead, on the premise that MZ twins are necessarily same-sex and about half of all DZ twins are same-sex we investigate our hypothesis restricting the sample to include only same-sex twins. If our results were driven by genetic predispositions then we should find weaker associations in the same-sex sample. The methods and data used to conduct the robustness checks are discussed alongside the results. The rest of this section elaborates the specification used in the two main approaches to testing for twin randomness.

Across Mothers: To test the null that twin births are "as good as random", we estimate conditional regressions of the form:

$$twin_{bjy} = \gamma_0 + \gamma_1 Health_{bjy} + \mu_b + \lambda_y + \varepsilon_{bjy}. \tag{1}$$

Here, $twin$ is an indicator of whether a birth of order b born to woman j at age y is a twin. We control for fixed effects for mother's age and parity, as these are known to influence the probability of twin birth. Where births are observed over multiple years, races or geographic areas, we include the relevant fixed effects. Under the null, the coefficients on maternal health variables $Health_{bjy}$

should not be statistically distinguishable from zero. This is equivalent to a test of (conditional) balance of characteristics of 'treated' (with twins) and 'control' (without twins) mothers. Standard errors are clustered at the level of the mother.

For ease of exposition, we maintain subscript y for the woman's age at birth but most of the health indicators are measured before pregnancy to avoid the potential concern of reverse causality, i.e. that twin births cause greater depletion of the mother's health than singleton births, or encourage women to adopt different behaviours. These include pre-pregnancy measures of smoking, diabetes, hypertension, obesity, height, kidney disease and asthma. Measures of prenatal or medical care are constructed as community-level measues of availability. In a specific case we discuss below we use an exogenous measure of environmental stress in pregnancy. We also show results for some variables measured in pregnancy—smoking, alcohol, drugs, diet—and for one measure (BMI in developing country data) measured after birth. We flag these variables so that their coefficients can be interpreted with this caveat in mind.[5] Importantly, if we dropped all of the flagged variables, we would still have a fairly compelling breadth of evidence. We add controls for education and, where available, wealth, to allow for the fact that education may motivate and wealth may facilitate health-seeking behaviours (Kenkel, 1991; Lleras-Muney and Cutler, 2010). This will confirm that

---

[5]The concern is that these variables may respond to a woman's knowledge that she is carrying twins. If the response is such as to accentuate the relationships of interest, for instance, if she smokes more, then failing to account for this would lead us to under-estimate the relationships of interest. However, if instead she increases her attendance of antenatal care and this more than offsets the resource stress of carrying twins, it is possible that we over-estimate the relationship. BMI in one data set is only available after birth. If twin births deplete the mother more, then twin mothers will record lower BMI, and accounting for this would only strengthen our contention.

the indicators in $Health$ are not simply proxying for socio-economic status. As discussed above, we will present additional specifications including woman fixed effects in the model and restricting to same-sex twins.

Pre-Twin Balance: We perform an alternative test that exploits pre-determined birth outcomes within-mothers. This essentially involves testing whether women who produce twins had, on average, healthier births *before* the twin birth, as this would be a measure of pre-determined maternal health. For each $n = \{2, 3, 4\}$ we estimate:

$$Prior\,Death_{b<n,jy} = \alpha_0 + \alpha_1 Twin_{b=n,jy} + \lambda_y + \nu_{jby}, \tag{2}$$

where we restrict the sample to prior birth outcomes of mother j who were fully exposed to the risk of death before birth order $b < n$. Thus, for $n = 2$, the independent variable $Twin$ takes the value of one if the mother gives birth to twins on her second birth, and zero if she gives birth to a singleton on her second birth. We generalize this to higher birth orders. $Prior\,Death$ refers to the proportion of pre-twin births of a mother which have survived and, for instance, for $n = 2$, this is the survival status of the first birth. When we use the US data, this refers to foetal survival and when we use the Demographic and Health Survey (DHS) data this refers to survival from birth through to 12 months of age. However we also show results for size at birth, a less extreme measure of child health than mortality.[6] If women who give birth to twins are selectively healthy we will observe $\alpha_1 < 0$. Maternal age fixed effects are included. In Appendix B.1, we discuss issues relating to the

---

[6]Infant mortality is widely used as a marker of health and it has the advantage that it is largely predetermined with respect to the following birth (given gestation is about 9 months), and to ensure this we remove children born less than a year after their older sibling. Similarly, miscarriage rates have been shown to respond to maternal condition, and are high, even in developed country settings.

measurement of maternal health and miscarriage data.

3   Data

Not all birth records contain indices of maternal health or health-related behaviours. To estimate equation 1 we sought data that did and that were representative and, given the relative rarity of twins, large. Data sets fulfilling these criteria include administrative birth data from the US, Spain and Sweden, and household survey data from Chile, the United Kingdom, and 68 developing countries (the DHS) for different sets of years. Details of temporal and geographic coverage, and summary statistics for each data set are provided in online data Appendix B.2. Together, these data sets include 17 million births through years 1972 to 2013. We consistently restrict the sample to women aged 18-49 years old, and exclude triplets and higher order multiple births. We take advantage of US Vital Statistics data from 2009 to 2013 that identify ART use by birth, removing the approximately 1.6% of births that were ART-assisted.[7] For the developing country sample, on the premise that ART was not available prior to 1990, we split the birth data into pre- and post-1990 samples.

Equation 2 is estimated using only the DHS and the US vital statistics files. The DHS has the complete fertility history including the survival status and birth weight of all children preceding each twin or singleton birth and the US birth certificate data allows us to infer earlier miscarriages for every mother as the difference between total reported births and live births. The miscarriage data are discussed further in Section 4.2.

---

[7]The data since 2009 also include a range of new measures of maternal morbidity and behaviours.

4   Results

4.1   Twin Births and Maternal Condition

In Table 2 we present estimates of equation 1 for several countries using multiple indicators of maternal health. We find broadly consistent results across indicators and across samples. In online Appendix C we provide additional discussion of the stability of the general result across countries and levels of economic development. All independent variables in Table 2 are standardised as Z-scores so that the estimates can be cast as the effects of increasing by 1 standard deviation (sd) the independent variable of interest. Unstandardised results are presented in Appendix Table A1.

We find that the probability of twin birth is significantly positively influenced by the following indicators of maternal health included independently: not underweight, tall[8], more educated, having greater access to medical or antenatal care, not having smoked before pregnancy, not having any of a range of morbidities prior to conception (obesity, diabetes, hypertension, asthma, kidney disease), and averting risky behaviours in pregnancy (smoking, alcohol, drugs, unhealthy diet). The effects are sizeable, with a 1 sd improvement in the indicator tending to increase the likelihood of twinning by 6-12% in most cases, relative to a mean of about 2.7% in the (global) sample. There are smaller effects from fresh fruit consumption and larger effects from height. We shall see when

---

[8]Height is the indicator of health most widely measured in birth and demographic data and several studies show that it responds to infection and nutritional scarcity in the growing years, eg individuals exposed to famine and war have been shown to have lower stature in adulthood, (Silventoinen, 2003; Bozzoli et al., 2009; Akresh et al., 2012). Previous research has shown widespread associations of short stature among mothers with the risk of low birth weight and infant mortality among their children (Bhalotra and Rawlings, 2013).

we present the pre-twin survival test results below that these effect sizes are broadly comparable to the difference in US data of about 7% in rates of miscarriage of first births between mothers who go on to have twins at second birth, and mothers who do not. This similarity of orders of magnitude contributes plausibility to our argument that miscarriage is a mechanism. We directly test this mechanism in section 4.2.

Using all available measures of health for each country, we also calculated a factor index of maternal health (as in Biroli (2016)); see Appendix D. Mothers of twins consistently have a higher score than mothers of singletons but, as the variables available for each country are different, the scores are not comparable across countries. Statistical significance of these health indicators is robust to running regressions which condition on all available indicators of the mother's health and, importantly, education (Appendix Tables A2-A3). Our results all hold after correcting test statistics for large sample sizes that increase the likelihood of rejecting a null, following Deaton (1997), see Appendix Table A4. First we will elaborate our findings by country. Then we present results from alternative approaches, and the robustness checks concerned with the role of genetic traits.

Estimates for the USA We pool all non-ART births in the United States during the years 2009 to 2013. We estimate that a 1 sd increase in rates of smoking before pregnancy is associated with a 0.11 percentage point (pp) lower chance of a twin birth which is about 5.5% of the mean rate of twinning.[9] Diabetes and hypertension prior to pregnancy have standardized effects of 0.2 to

---

[9]Effects of smoking during pregnancy are larger, in the range of 0.20-0.24 pp, with smoking in the third trimester imposes the largest reduction, consistent with evidence that adverse effects of smoking on birth weight are largest in the third trimester – see Bernstein et al. (2005) and a similar pattern estimated on our data in Appendix Table A5.

0.3 pp while being obese or underweight prior to pregnancy has smaller effects of 0.04 and 0.16 pp respectively. Height and education have larger standardized effects, of 0.61 and 0.8 pp respectively. In Appendix Table A6 we remove potential outliers from the sample of mothers when considering height and the results are nearly entirely unchanged. Estimates for women using ART are presented in Table A7 and are, with the exception of being underweight, larger and statistically significant for every indicator, underlining the additional sensitivity of birth outcomes in this group.

Estimates for Sweden, Avon (UK) and Chile Analysis of birth registers from Sweden for 1993-2012 indicates strikingly similar standardised effect sizes for smoking, diabetes, height and being underweight to those for US women. There are however some differences: the standardized co-efficient on obesity in Sweden is about three times as large, while the coefficient on hypertension is only half as large. The Swedish data additionally record asthma prior to conception, which we estimate reduces the risk of twin births by 0.015 pp. Survey data from Avon county UK 1991-1992 and Chile 2006-2009 again exhibit patterns similar to those identified for Sweden and the USA for anthropometric indicators of health, risky behaviours and pre-pregnancy illnesses. For instance, for the UK, estimates for being underweight, obese or smoking before pregnancy are all very similar to the corresponding estimates for the US. However the standardized impact of hypertension before pregnancy is twice as large, and the associations with diabetes, height and education are smaller. The UK data contain unique information on eating healthily during pregnancy and our estimates indicate that the standardised effect of this is a 0.54 pp increase in the likelihood of having twins, which is the single largest coefficient among variables available for the UK. The coefficients in the Chilean data for being underweight and for smoking, drugs and alcohol consumption during pregnancy lie between 0.16 and 0.33 pp, broadly similar to the coefficients for other countries, and the coefficient on obesity is considerably larger (0.26). Chile is the only country in our sample for

which we have information on drug use during pregnancy and the standardized effect for this is similar to that for (frequent) alcohol consumption in pregnancy.

Estimates for Developing Countries In the sample that pools data for 68 developing countries for 1972-2012, we observe height, weight, body mass index, and local availability of prenatal care and access to medical professionals. Reproductive health service coverage is far from universal in low-income countries, although this is a leading global health priority.[10] After adjusting for demographic covariates as for the other samples, we observe again that taller and heavier women are more likely to twin. This is true even in the pre-ART period (see Table A8). The effects of height, underweight and education are all smaller than in richer countries, while the effects of obesity are larger than in all countries other than Chile.[11] We estimate that a 1 sd increase in availability of doctors or nurses is associated with a 0.092 pp and 0.06 pp increase in the likelihood of twins respectively.

Quasi-experimental variation in a negative intrauterine shock: Spain Using the methodology and data described in Quintana-Domeque and Ródenas-Serrano (2017), we estimated the impact of ETA bombing as a plausibly exogenous negative intrauterine shock which may cause foetal stress, a proxy for maternal health in pregnancy. We find that an additional bomb casualty in the province of residence of a pregnant woman decreases the likelihood that she will have a twin birth by 0.01% and 0.012%; see Table 3. This effect is larger and only statistically significant during the second and third trimesters, similar to the effects of smoking by trimester documented in Table 2.[12]

---

[10]These variables are all measured as the rate of healthcare access in the mother's cluster of residence since we are interested in availability rather than use to avoid the concern that mothers conceiving twins may be more likely to actively seek birth attendance.

[11]Recall these are standardized effects; unstandardized effects are in the Appendix.

[12]Quintana-Domeque and Ródenas-Serrano (2017) find that the same exposure reduces average

Survival of pre-twins as a marker of mother's health Here we discuss the alternative test of the quasi-randomness of twin births. Estimates of equation 2 are in Table 4. In the developing country sample, mothers who went on to have third- and fourth-born twins had an infant mortality rate 1.3-1.7 percentage points lower among their prior births than women who had singletons at the same birth order. This is a natural measure of maternal health, capturing a woman's ability to produce surviving children, which is exactly what we hypothesize is challenged by carrying twins. We used birth size as a measure of child health that is less extreme than mortality. We used the DHS again, as it allows us to observe all children ordered within mother, and we find that earlier births of women who later have a twin birth are less likely to be small at birth than the corresponding births of women who have only singleton children (see Appendix Table A9).

Similarly, in the US population, we observe that women who have twins are less likely to have suffered a miscarriage prior to the twin birth. Mothers who give birth to twins at second birth are 0.7 percentage points less likely to have suffered a miscarriage of their first conception, which is 6.7% of the baseline rate for this group. The rate of miscarriages in the population of all women who gave birth was approximately 10%. Parity-specific estimates and means are in Table 4.

Specification check including woman fixed effects So as to control for any genetic characteristics of the mother, we sought data that follow women over time, recording multiple births per woman as well as *time-varying* measures of maternal health. Such data are scarce, but the National Longitudinal Survey of Young Women (NLSY) meet these requirements. A sample of 5,159 women aged 14 and 24 in 1968 is followed until 1999, when the youngest are aged 45. The health vari-

birth weight by approximately 0.3 grams (trimester 1), and increases the likelihood of low birth weight by 0.14%. Placebo tests in support of their methodology including examining the impact of bombs *post*-birth on birth outcomes are presented in their paper.

ables measured consistently through this period are whether the mother has any physical limitation which restricts her ability to work, whether she smoked prior to the pregnancy, and whether she has had a prior cancer diagnosis. More information on the data structure, and summary statistics is in Appendix E. We estimate the probability that a birth is a twin as a function of these indicators of maternal health conditional upon mother fixed effects and controlling also for a quadratic in family income, mother's age, birth order and year of birth fixed effects. Results in Appendix Table A17. We find large statistically significant negative effects of smoking and cancer on the probability of having a twin birth, and no significant impact of health limiting work.

Specification check using monozygotic twins The risk of giving birth to dizygotic twins (DZ) is elevated among women with high levels of the follicle stimulating hormone (FSH), which is often more prevalent among taller and heavier women (Li et al., 2003; Hall, 2003; Hoekstra et al., 2008). Since dizygotic twins constitute about two-thirds of all twins, this could in principle contribute to explaining the associations we document with height and BMI (note that the biomedical literature has not documented these associations in any population level data, let alone across countries, time and indicators). Although, as discussed, genetic predispositions cannot explain our finding that health behaviours or aspects of the health environment (stress or prenatal care availability) predict twinning, we investigate this further by exploiting the fact that MZ twins are necessarily same-sex (and about half of DZ twins are same-sex) and repeat the analysis removing mixed-sex twins from the data.[13] Results are in Appendix Table A10. We continue to find significant associations between

---

[13]Note that this also addresses the elevation of twin birth rates among ART users for the DHS where ART use is not recorded, because most if not all of ART-generated twins are DZ. We implement this test using DHS data only, as in other administrative data sets, twins are not matched with their siblings to infer whether they are of the same sex.

proxies for maternal health and the chances of a twin vs a singleton birth and the coefficients are not significantly different from those that obtain in the full sample.

4.2 Mechanisms of Twin Selection

We consider three alternative hypotheses for why maternal health may influence the probability of twinning, which relate to conception, gestation and maternal survival. First, healthier mothers may be more likely to *conceive* twins on account of an underlying genetic or biological process. Second, conditional upon conceiving twins, healthier mothers may be more likely to take them to term. Third, conditional on conceiving twins and taking them to term, healthier mothers may be more likely to survive the birth, and hence appear in the available data.

Either of the first two processes is sufficient to violate the "as good as random" assumption insofar as they imply that observing twins will depend upon possibly unmeasured maternal behaviours and characteristics. Since taller and heavier women, and active smokers have higher levels of the FSH hormone associated with multiple births (Li et al., 2003; Hall, 2003; Hoekstra et al., 2008; Cramer et al., 1994), conception of twins may not be random. We cannot directly test the conception hypothesis since the required data are unavailable but we now provide tests of the other two hypotheses and indicate the manner in which non-random conception will influence interpretation of our results.

Selective foetal death The gestation hypothesis is that carrying twins to term is more demanding than carrying singletons to term, and so stressors of maternal health will lead to selective miscarriage of twins. It has been documented that the biological demands of twin pregnancies are higher than the demands of non-twin pregnancies (Shinagawa et al., 2005) and also that, in general, healthier mothers are less likely to miscarry (García-Enguídanosa et al., 2002). What we contribute here

17

is to test the natural intersection of these hypotheses, and estimate the extent to which miscarriage is more frequent among less-healthy women carrying twins. The estimated equation is:

$$FoetalDeath_{ijt} = \gamma_0 + \gamma_1 Twin_{ijt} + \gamma_2 Health_{jt} + \gamma_3 Twin \times Health_{ijt} + \lambda_t + \phi_y + \mu_b + u_{ijt}. \quad (3)$$

$FoetalDeath_{ijt}$ is a binary variable (multiplied by 1,000) indicating whether a birth was taken to term (coded as 0) or resulted in a miscarriage (coded as 1), i indicates a conception leading to birth or foetal death, j a mother, and t is year. $Health$ is an indicator of the mother's health, $Twin$ is an indicator for whether the conception is a twin or a singleton and, as before, fixed effects for year (λ_t), birth order (μ_b), and mother's age (ϕ_y) are included. The coefficient of interest γ_3 is the *differential* effect of the variable $Health_{jt}$ on twin conceptions.

Birth registers often do not include maternal health indicators and if they do it is unusual that they also also include information on foetal deaths, but the US Vital Statistics data do.[14] We pooled all births and foetal deaths recorded during 1999-2002. We stopped in this year because, from 2003, a considerable re-definition of birth certificate data meant that foetal death and birth data did not share similar controls. Prior to 2002 however we are able to observe for all states whether a mother smokes or drinks during pregnancy, whether she suffered from anemia prior to pregnancy, and her educational level. The results, using the US birth certificate and fetal death data, are in Table 5. In panel A we first document the difference in the risk of foetal death for twin relative to singleton

---

[14]We would have liked to replicate this analysis in other data sets. However, we are not aware of other data that has all the details necessary to run such a test, in particular maternal health outcomes, births, miscarriages and an indicator of whether miscarried births were twin or singleton. For instance, in DHS data miscarriages are recorded in certain surveys – such as for Nepal – but there is no record of whether these they are for single or twin pregnancies.

conceptions. The evidence confirms previous research showing that the spontaneous abortion rate among twins (at 1 in 8 conceptions), is about three times that among singletons (Boklage, 1990). In panel B, we test how maternal health indicators modify this differential risk. We can consistently reject that the interaction term γ_3 is zero. In other words, twin foetal survival is more sensitive to mother's health than singleton survival. For example, a 1 standard deviation increase in rates of smoking whilst carrying a singleton elevates the risk of miscarriage by 1.39 foetal deaths per 1,000 live births. The corresponding risk elevation among mothers pregnant with twins is an increase of 2.55 foetal deaths, almost twice the risk. Alcohol consumption is similarly almost twice as risky for women carrying twins, and the risks associated with anemia are about three times as high. We also see that a college education, which is a predictor of healthy behaviour, modifies the difference in miscarriage probabilities more than three times as much when the mother is carrying twins than when she is carrying a singleton. Now it may be that *one of two twins miscarries*. In such cases, if the survivor is recorded as a singleton birth then we will tend to under-estimate the importance of maternal condition. In other words, our contention holds *a fortiori*.

Overall, these results establish a plausible mechanism for the associations that we document in Tables 2–4. Here we have modeled miscarriage conditional upon the conception being twin or singleton. If in fact maternal health also raises the chances of a twin conception, then this will reinforce our contention. If, instead, maternal health is for some undocumented reason negatively associated with twin conception, then our findings hold despite this and are conservative.

Trivers and Willard (1973) made an argument similar to ours but pertaining to the distribution of sons across women (Trivers and Willard, 1973; Almond and Edlund, 2007). They observed that since the male foetus is more vulnerable to adverse health conditions (Waldron, 1983), sons are more likely to be born of healthy mothers. As for twins, so for sons, selective miscarriage is

the suggested mechanism. Intersecting our hypothesis with theirs, we investigated whether males are under-represented among twins, other things equal. We used the large data sets in Table 2 (US, Sweden and the developing country data). We find that twin births are approximately 0.1-0.3 percentage points more likely to be female ($p<0.001$). This affords a further test of our hypothesis and a validation of the Trivers-Willard hypothesis (refer to Appendix Table A12).[15]

Our findings suggest that twin birth is a marker of foetal health. Our findings, that range across indicators and countries, highlight the relevance of maternal health for foetal health. Recent research demonstrating long run socio-economic returns to investing in foetal and infant health, improving the pre-school environment and raising parenting quality has stimulated policy interventions across the world that are motivated to enhance the potential for nurture to lift up the trajectories of children, especially when born into disadvantaged circumstances (Heckman et al., 2010; Almond and Currie, 2011; Carneiro et al., 2015). Our results point to the significance of, for instance, nutrition, stress and prenatal care for mothers in achieving these goals.

Selective maternal survival A potential concern is that the less-healthy women among those who delivered twins died in childbirth, and data sets like the DHS that obtain birth histories from mothers will not contain those women. In such cases our findings could arise from selective maternal survival. This concern does not apply to the administrative US and Swedish data where

---

[15]We found an older biological literature which recognizes that males are under-represented among twins, and even more under-represented among triplets (James, 1975; Bulmer, 1970), but this literature does not explicitly link in with Trivers-Willard. When interacting twin births by maternal characteristics in Table A12 most coefficients are not significantly different by the gender of the twins. However, two coefficients are significantly larger (more negative) for boys, consistent with the male foetus being more sensitive to foetal health.

all births are recorded and where we see clear associations of twinning and maternal health so it cannot be the only explanation of those findings. Similarly, in the UK and Chile data sets, the survey design ensures that representative coverage is not affected by maternal death.[16] The lifetime risk of a maternal death is 1 in 41 in low income countries as compared with 1 in 3300 in high income countries. If twins only spuriously appear to be born of healthier mothers due to selective maternal death, then as mothers become more likely to survive childbirth (ie as maternal mortality declines), the associations should dissipate. The fact that they do not also undermines the relevance of selection.

We assess the magnitude of selection bias in our DHS estimates, following Alderman et al. (2011). We simulate the presence of the women who died and test whether correcting for maternal survival selection causes the association of twin births and maternal health to disappear.

A data challenge is that we do not observe the health of women who died in childbirth, indeed, the original problem is that we do not observe these women at all. We address this by using the maternal mortality status of all sisters of every female respondent.[17] We assume that the respondent's health (indicated by height and BMI) proxies the health of her sisters, and validate this (Figure A1).[18] We put our results to the harshest test by assuming that less-healthy women who died in childbirth were all carrying twins, and more healthy women who died in childbirth were not carry-

---

[16]In data from the UK, women were prospectively enrolled when pregnant entirely before exposure to considerable maternal mortality risk, and children were subsequently followed over their lives. In the data from Chile, a representative sample was chosen *after* birth, however the sampling unit was at the level of the child, rather than the mother, so children would be represented even in cases where their mother was no longer alive.

[17]Most DHS countries are in Africa where fertility and maternal mortality are high.

[18]Maternal mortality is significantly higher among sisters of women with lower height or BMI,

ing twins, and the results stand up to this, see Table 6. We test sensitivity of the adjusted estimates to a range of different binary distinctions of healthy vs less-healthy. Overall, these results establish that maternal mortality selection does not drive the DHS results.

Conclusion and Discussion

Twin births are not random. We show that mothers who have twin births are healthier *prior* to the occurrence of the twin birth. The findings in this paper have implications for identification strategies in economics and a number of other fields of research, and they extend the existing social science and bio-medical literature on twinning. Here we delineate these contributions, using a list format for clarity.

(i) The biomedical literature has identified an association of twinning with the height, weight and smoking status of the mother, and attributed this to hormonal variation. This is the first study to demonstrate that these associations hold in representative population-level data in several richer and poorer countries and across several years. We also show that these associations hold conditional not only on age and parity (known predictors) but also upon the mother's socio-economic status and on a range of other indicators of her health.

(ii) This is the first paper to demonstrate associations of twinning with other indicators of maternal health. These include a range of (pre-pregnancy and pregnancy) morbidities, the health of her lower order births, the mother's health-related behaviours (before and during pregnancy), availability of reproductive health services, and indicators of the mother's exposure to environmental

---

conditional upon country and year fixed effects, a quadratic in mother's age and age at first birth. Sisters of women shorter than the mean height of 155.5cm are considerably more likely to have suffered maternal death, and there is a sharper gradient for women shorter than 145cm.

stress in pregnancy. The last three are clearly not genetic or hormonal associations. We neverthe-less show associations of maternal health and twinning conditional upon woman fixed effects that purge genetic differences between women.

(iii) Since it is known that twins are more likely among ART-assisted births and that ART-users tend to be more educated, we show that associations of maternal health with twinning hold in ART-purged and pre-ART data samples. We also show the first systematic evidence that the education of the mother is positively associated with twinning in these samples, consistent with educated women being more likely to engage in health seeking behaviours.

(iv) Our findings indicate no clear tendency for the association of maternal health with twinning to dissipate with economic development. Although intrinsic maternal health and access to public health services tend to improve with economic development, it is unclear that all relevant indicators (hypertension, obesity, diabetes) improve, and differences between rich and poor countries in age, parity and race will also modify this relationship.

(v) We are able to demonstrate that maternal health determines foetal selection conditional upon conception of twins. The bio-medical literature has discussed hormonal (FSH) predictors; our hypothesis that it is selectively healthy women who are able to mount the challenge of carrying twins to birth is new.

In the economics literature, the validity of several studies investigating the hypothesis that fertility has a causal effect on investments in children, or on women's labour supply, rests upon the assumption that twin births are random (at least conditional on age, parity and education). Twin births are used as an instrument because OLS estimates tend to be biased upward on account of negative selection of women into fertility. Our findings suggest that twin-IV estimates will tend to be biased downward on account of positive selection into twin birth. This is important because re-

cent prominent studies cited in the Introduction find that the trade-off is frequently not statistically different from zero and, in principle, this could be explained by a downward bias in the estimates.[19]

Educational attainment has risen considerably while completed and desired fertility have fallen sharply over the past 50 years (see *eg* Hanushek (1992)). It is of considerable relevance to researchers and to policy makers to determine whether these trends contain a causal component. Similarly, the fertility-work trade-off for women is topical again as educational attainments of women are over-taking those of men and transforming the work-family balance, with consequences for women's autonomy, marital stability and child outcomes (Newman and Olivetti, 2016; Lundberg et al., 2016).

References

R. Akresh, S. Bhalotra, M. Leone, and U. Osili. War and Stature: Growing Up During the Nigerian Civil War. *American Economic Review (Papers & Proceedings)*, 102(3):273–77, 2012.

H. Alderman, M. Lokshin, and S. Radyakin. Tall claims: Mortality selection and the height of children. Policy Research Working Paper 5846, The World Bank, Oct 2011.

D. Almond and J. Currie. Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives*, 25(3):153–172, 2011.

D. Almond and L. Edlund. Trivers–Willard at birth and one year: evidence from US natality data

---

[19]For the estimates to be biased, we require not only that the probability of a twin birth is non-random and associated with potentially unobservable maternal health, but also that maternal health is correlated with the propensity to invest in children or to participate in the labour force, as the case may be. There is evidence of the latter.

1983–2001. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1624): 2491–2496, 2007.

J. Angrist, V. Lavy, and A. Schlosser. Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, 28(4):pp. 773–824, 2010.

J. D. Angrist and W. N. Evans. Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *American Economic Review*, 88(3):450–77, June 1998.

O. Åslund and H. Grönqvist. Family size and child outcomes: Is there really no trade-off? *Labour Economics*, 17(1):130–39, 2010.

G. S. Becker and H. G. Lewis. On the interaction between the quantity and quality of children. *Journal of Political Economy*, 81(2):S279–88, Part II, 1973.

I. M. Bernstein, J. A. Mongeon, G. J. Badger, L. Solomon, S. H. Heil, and S. T. Higgins. Maternal smoking and its association with birth weight. *Obstetrics and Gynecology*, 106(5):986–991, 2005.

S. Bhalotra and S. Rawlings. Gradients of Intergenerational Transmission of Health in Developing Countries. *The Review of Economics and Statistics*, 95(2):660–672, 2013.

P. Biroli. Health and Skill Formation in Early Childhood. Working Paper 17, University of Zurich, UBS International Center of Economics in Society, Nov. 2016.

S. E. Black, P. J. Devereux, and K. G. Salvanes. The more the merrier? the effect of family size and birth order on children's education. *The Quarterly Journal of Economics*, 120(2):669–700, 2005.

S. E. Black, P. J. Devereux, and K. G. Salvanes. Small Family, Smart Family? Family Size and the IQ Scores of Young Men. *Journal of Human Resources*, 45(1):33–58, 2010.

S. E. Black, P. J. Devereux, and K. G. Salvanes. Does Grief Transfer across Generations? Bereavements during Pregnancy and Child Outcomes. *American Economic Journal: Applied Economics*, 8(1):193–223, January 2016.

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

C. E. Boklage. Survival probability of human conceptions from fertilization to term. *International Journal of Fertility*, 35(2):79–94, 1990.

C. Bozzoli, A. Deaton, and C. Quintana-Domeque. Adult height and childhood disease. *Demography*, 46(4):647–669, November 2009.

S. G. Bronars and J. Grogger. The economic consequences of unwed motherhood: Using twin births as a natural experiment. *The American Economic Review*, 84(5):1141–1156, 1994.

M. G. Bulmer. *The Biology of Twinning in Man*. Oxford Clarendon Press, Oxford, UK, 1970.

J. Cáceres-Delpiano. The impacts of family size on investment in child quality. *Journal of Human Resources*, 41(4):738–754, 2006.

J. Cáceres-Delpiano. Can We Still Learn Something From the Relationship Between Fertility and Mother's Employment? Evidence From Developing Countries. *Demography*, 49(1):151–174, February 2012.

D. M. Campbell. Epidemiology of twinning. *Current Obstetrics & Gynaecology*, 8(1):126–134,

1998.

D. M. Campbell, A. J. Campbell, and I. MacGillivary. Maternal Characteristics of Women Having Twin Pregnancies. *Journal of Biosocial Science*, 6(1):463–470, 1974.

P. Carneiro, K. Løken, and K. G. Salvanes. A flying start? maternity leave benefits and long-run outcomes of children. *Journal of Political Economy*, 123(2):365–412, 2015.

K. Chay and M. Greenstone. The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession. *The Quarterly Journal of Economics*, 118(3):1121–1167, 2003.

D. Clarke. Children And Their Parents: A Review Of Fertility And Causality. *Journal of Economic Surveys*, 32(2):518–540, April 2018.

D. W. Cramer, R. L. Barbieri, H. Xu, and J. K. Reichardt. Determinants of basal follicle-stimulating hormone levels in premenopausal women. *The Journal of Clinical Endocrinology & Metabolism*, 79(4):1105–1109, 1994.

J. Currie and E. Moretti. Biology as Destiny? Short- and Long-Run Determinants of Intergenerational Transmission of Birth Weight. *Journal of Labor Economics*, 25(2):231–264, 2007.

A. Deaton. *The Analysis of Household Surveys – A Microeconometric Approach to Development Policy*. The Johns Hopkins University Press, 1997.

E. Fitzsimons and B. Malde. Empirically probing the quantity-quality model. *Journal of Population Economics*, 27(1):33–68, Jan 2014.

A. García-Enguídanosa, M. Calleb, J. Valeroc, S. Lunaa, and V. Domínguez-Roja. Risk factors in

miscarriage: a review. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 102(2):111–119, May 2002.

J. G. Hall. Twinning. *The Lancet*, 362(9385):735–743, August 2003.

E. A. Hanushek. The trade-off between child quantity and quality. *Journal of Political Economy*, 100(1):84–117, February 1992.

J. J. Heckman, S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz. The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, 94(1):114–128, 2010.

C. Hoekstra, Z. Z. Zhao, C. B. Lambalk, G. Willemsen, N. G. Martin, D. I. Boomsma, and G. W. Montgomery. Dizygotic twinning. *Human Reproduction Update*, 14(1):37–47, 2008.

J. P. Jacobsen, J. W. P. III, and J. L. Rosenbloom. 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):449–474, 1999.

W. H. James. Sex ratio in twin births. *Annals of Human Biology*, 2(4):365–378, 1975.

D. S. Kenkel. Health Behavior, Health Knowledge, and Schooling. *Journal of Political Economy*, 99(2):287–305, 1991.

H. Li, J. Zhang, and Y. Zhu. The quantity-quality trade-off of children in a developing country: Identification using Chinese twins. *Demography*, 45:223–243, 2008.

Z. Li, J. Gindler, and H. Wang. Folic acid supplements during early pregnancy and likelihood of multiple births: a population-based cohort study. *The Lancet*, 361:380–84, 2003.

A. Lleras-Muney and D. Cutler. Understanding Differences in Health behaviors by Education.

Journal of Health Economics, 29(1):1–28, 2010.

S. Lundberg, R. A. Pollak, and J. Stearns. Family Inequality: Diverging Patterns in Marriage, Cohabitation, and Childbearing. *Journal of Economic Perspectives*, 30(2):79–102, 2016.

P. Lundborg, E. Plug, and A. W. Rasmussen. Can Women Have Children and a Career? IV Evidence from IVF Treatments. *American Economic Review*, 107(6):1611–37, June 2017.

L. J. Marteleto and L. R. de Souza. The Changing Impact of Family Size on Adolescents' Schooling: Assessing the Exogenous Variation in Fertility Using Twins in Brazil. *Demography*, 49(4):1453–1477, 2012.

B. Mazumder and Z. Seeskin. Breakfast skipping, extreme commutes and the sex composition at birth. *Biodemography and Social Biology*, 61(2):187–208, 2015.

W. J. Meulemans, C. M. Lewis, D. I. Boomsma, C. A. Derom, H. Van den Berghe, J. F. Orlebeke, R. F. Vlietinck, and R. M. Derom. Genetic modelling of dizygotic twinning in pedigrees of spontaneous dizygotic twins. *American Journal of Medical Genetics*, 61(3):258–263, 1996. ISSN 1096-8628.

M. Mogstad and M. Wiswall. Testing the Quantity-Quality Model of Fertility: Linearity, Marginal Effects, and Total Effects. *Quantitative Economics*, 7(1):157–192, 2016.

A. F. Newman and C. Olivetti. Career Women and the Durability of Marriage. mimeo, Boston University, 2016.

T. J. C. Polderman, B. Benyamin, C. A. de Leeux, P. F. Sullivan, A. van Bochoven, P. M. Visscher, and D. Posthuma. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, 47(7):702–709, May 2015.

V. Ponczek and A. P. Souza. New Evidence of the Causal Effect of Family Size on Child Quality in a Developing Country. *Journal of Human Resources*, 47(1):64–106, 2012.

C. Quintana-Domeque and P. Ródenas-Serrano. The Hidden Costs of Terrorism: The Effects on Health at Birth. *Journal of Health Economics*, 56:47–60, 2017.

M. R. Rosenzweig and K. I. Wolpin. Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica*, 48(1):227–40, January 1980a.

M. R. Rosenzweig and K. I. Wolpin. Life-cycle labor supply and fertility: Causal inferences from household models. *Journal of Political Economy*, 88(2):pp. 328–348, 1980b.

M. R. Rosenzweig and K. I. Wolpin. Natural "Natural Experiments" in Economics. *Journal of Economic Literature*, 38(4):827–874, December 2000.

M. R. Rosenzweig and J. Zhang. Do population control policies induce more human capital investment? twins, birth weight and China's one-child policy. *Review of Economic Studies*, 76(3): 1149–1174, 2009.

S. Shinagawa, S. Suzuki, H. Chihara, Y. Otsubo, T. Takeshita, and T. Araki. Maternal basal metabolic rate in twin pregnancy. *Gynecologic and Obstetric Investigation*, 60(3):145–48, 2005.

K. Silventoinen. Determinants of variation in adult body height. *Journal of Biosocial Science*, 35 (2):263–285, April 2003.

E. L. Thorndike. Measurement of Twins. *The Journal of Philosophy, Psychology and Scientific Methods*, 2(2):547–553, Sep 1905.

R. L. Trivers and D. E. Willard. Natural selection of parental ability to vary the sex ratio of offspring.

Science, 179(4068):90–92, 1973.

S. Vitthala, T. A. Gelbaya, D. R. Brison, C. T. Fitzgerald, and L. G. Nardo. The risk of monozygotic twins after assisted reproductive technology: a systematic review and meta-analysis. *Human Reproduction Update*, 15(1):45–55, Jan-Feb 2009.

I. Waldron. The role of genetic and biological factors in sex differences in mortality. In A. D. Lopez and L. T. Ruzicka, editors, *Sex differentials in mortality: trends, determinants and consequences*, (Department of Demography Miscellaneous Series, chapter 4, pages 141–164. Australian National University, Department of Demography, 1983.

Table 1: The Quantity–Quality and Fertility–FLFP Tradeoffs: Estimates using the Twin Instrument

Authors	Data/Context	Twin Use	Estimates		Maternal Controls
			OLS	IV	
Panel A: Quantity–Quality					
Rosenzweig and Wolpin (1980a)	India, rural survey. Outcome is standardised schooling	Twin ratio in OLS[a]	-2.483 (0.740)		None
Black et al. (2005)	Norway, administrative data.	IV	-0.060 (0.003)	-0.038 (0.047)	Age and education
			-0.076 (0.004)	-0.016 (0.044)	
			-0.059 (0.006)	-0.024 (0.059)	
Cáceres-Delpiano (2006)	US, census 5% file. Outcome is behind educational cohort	IV	0.011 (0.000)	0.002 (0.003)	Age, education and race
			0.017 (0.001)	0.010 (0.006)	
Li et al. (2008)	China, census 1% file. Outcome is educational enrollment	IV‡	-0.031 (0.001)	0.002 (0.009)	Age and education
			-0.038 (0.002)	-0.024 (0.014)	

Study	Data	RF[‡,b]	No Birthweight control	Birthweight control	Age
Rosenzweig and Zhang (2009)	China, twin survey.	RF[‡,b]	-0.307 (0.160)	-0.225 (0.172)	Age
Angrist et al. (2010)	Israel, census 20% file.	IV	-0.145 (0.005)	0.174 (0.166)	Age, place of birth, race
			-0.143 (0.005)	0.167 (0.117)	
Black et al. (2010)	Norway, administrative data. Outcome is IQ.	IV		-0.149 (0.052)	Age and education
				-0.170 (0.052)	
				-0.115 (0.080)	
Åslund and Grönqvist (2010)	Sweden, administrative data.	IV	-0.113 (0.004)	0.022 (0.048)	Age and education.
			-0.132 (0.006)	-0.043 (0.048)	
			-0.100 (0.009)	-0.042 (0.083)	
Ponczek and Souza (2012)	Brazil, census 10% file	IV (Girl)	-0.277 (0.015)	-0.372 (0.198)	Age and education
			-0.283 (0.015)	-0.634 (0.194)	
		IV (Boy)	-0.233 (0.010)	-0.137 (0.146)	
			-0.230 (0.010)	-0.060 (0.164)	
Marteleto and de Souza (2012)	Brazil, household survey	IV	-0.248 (0.003)	0.064 (0.076)	Age, education and family income
			-0.240 (0.003)	0.131 (0.055)	

Study	Data	Method	Estimate	Estimate	Controls
Mogstad and Wiswall (2016)	Norway, administrative data.	IV[c]		0.053 (0.050) -0.051 (0.053) -0.107 (0.059)	Age and education.

Panel B: Fertility and Female Labour Force Participation

Study	Data	Method	Estimate	Estimate	Controls
Rosenzweig and Wolpin (1980b)	US, pooled demographic surveys	RF[d]		-0.371 (0.212) Short-term estimate; 0.142 (0.102) Long-term estimate	None
Bronars and Grogger (1994)	US 1970 and 1980 5% Census	RF[d]		-0.036 (0.036) 1970 Census; -0.035 (0.017) 1980 Census	Age at first birth.
Angrist and Evans (1998)	US 1980 5% Census	IV	-0.176 (0.002)	-0.057 (0.011)	Age, age at first birth
Jacobsen et al. (1999)	US 1970 and 1980 5% Census	IV[e]		-0.021 (0.014) -0.025 (0.008)	Age at first birth cubic
Cáceres-Delpiano (2012)	Pooled demographic surveys, developing countries	IV	-0.014 (0.001) -0.010 (0.001) -0.009 (0.001)	-0.029 (0.012) -0.016 (0.012) -0.022 (0.012)	Age, education, literacy status, country dummies.

Estimates and standard errors reported in columns OLS and IV refer to main estimates from published articles using large samples of microdata. A comprehensive review is provided in Clarke (2018). Where multiple estimates are reported, unless otherwise indicated, the first line refers to the impact of twins at birth two, the second line the impact of twins at birth three, and the third line the impact of twins at birth four (if available). In panel A the estimates refer to the outcome variable "years of education" unless specified in column 2. In panel B, all outcomes are the mother's labour market participation. [a] Twin Ratio is the number of twin births divided by the number of pregnancies. [b] Coefficients reported are impact of second birth twins on non-twin first births. [c] Non-linear estimates are reported in paper. Here linear estimates are presented for comparison with other results. [d] Reduced form uses twins at first birth as independent variable. [e] First line reports estimates from 1970 census, second line reports 1980 census. ‡ Standard errors are calculated from reported *t*-statistics.

Table 2: Effects of Maternal Health on Twin Births

	Health Behaviours / Access			Health Stocks and Conditions		
Variable	Estimate	[95% CI]	Variable	Estimate	[95% CI]
Panel A: United States [N =13,646,236, % Twin = 2.84]					
Smoked Before Pregnancy	-0.108***	[-0.116,-0.100]	Height	0.612***	[0.604,0.620]
‡Smoked Trimester 1	-0.195***	[-0.203,-0.187]	Underweight	-0.156***	[-0.164,-0.148]
‡Smoked Trimester 2	-0.232***	[-0.240,-0.224]	Obese	0.042***	[0.032,0.052]
‡Smoked Trimester 3	-0.238***	[-0.246,-0.230]	Diabetes	-0.286***	[-0.296,-0.276]
Education	0.800***	[0.790,0.810]	Hypertension	-0.223***	[-0.233,-0.213]
Panel B: Sweden [N =1,240,621, % Twin = 2.55]					
‡Smoked (12 weeks)	-0.266***	[-0.301,-0.231]	Height	0.617***	[0.592,0.642]

‡Smoked (30-32 weeks)	-0.285***	[-0.312,-0.258]	Underweight	-0.140***	[-0.173,-0.107]
			Obese	-0.113***	[-0.137,-0.089]
			Asthma	-0.015*	[-0.033,0.003]
			Diabetes	-0.253***	[-0.278,-0.228]
			Kidney Disease	-0.079***	[-0.101,-0.057]
			Hypertension	-0.099***	[-0.121,-0.077]

Panel C: United Kingdom (Avon) [N =10,463, % Twin = 2.37]

‡Healthy Foods	0.538***	[0.256,0.820]	Height	0.399***	[0.115,0.683]
‡Fresh Fruit	0.019	[-0.281,0.319]	Underweight	-0.161	[-0.439,0.117]
‡Alcohol (Infrequently)	-0.099	[-0.373,0.175]	Obese	-0.046	[-0.322,0.230]
‡Alcohol (Frequently)	-0.358**	[-0.630,-0.086]	Diabetes	-0.056	[-0.328,0.216]

‡Passive Smoke	0.047	[-0.243,0.337]	Hypertension	-0.480***	[-0.752,-0.208]
‡Smoked during Pregnancy	-0.162	[-0.448,0.124]			
Education	0.416*	[-0.002,0.834]			

Panel D: Chile [N =14,050, % Twin = 2.55]

‡Smoked during Pregnancy	-0.327***	[-0.572,-0.082]	Underweight	-0.183*	[-0.399,0.033]
‡Drugs (Infrequently)	0.002	[-0.253,0.257]	Obese	-0.258***	[-0.446,-0.070]
‡Drugs (Frequently)	-0.161***	[-0.196,-0.126]			
‡Alcohol (Infrequently)	-0.072	[-0.362,0.218]			
‡Alcohol (Frequently)	-0.172***	[-0.213,-0.131]			
Education	0.529***	[0.200,0.858]			

Panel E: Developing Countries [N =2,050,795, % Twin = 2.07]

Doctor Availability	0.092***	[0.059,0.125]	Height	0.276***	[0.245,0.307]
Nurse Availability	0.060***	[0.029,0.091]	Underweight	-0.090***	[-0.115,-0.065]
Prenatal Care Availability	0.103***	[0.076,0.130]	Obese	0.059***	[0.028,0.090]
Education	0.141***	[0.110,0.172]		

‡: Conditions which are measured during pregnancy, and so may be behavioural responses to twins.

Each coefficient represents a separate regression of child's birth type (twin or singleton) on the mother's health behaviours and conditions. In each sample, all mothers aged 18-49 are included. Twins (dependent variable) is mutliplied by 100 and the independent variables are standardised as Z-scores so coefficients are interpreted as the percentage point change in twin births associated with a 1 standard deviation increase in the variable of interest. All models include fixed effects for age and birth order, and where possible, for wealth (panels A and D) and for gestation of the birth in weeks (panels A and B). Unstandardised and conditional results are included as online Appendix Tables A1 and A2. Results are robust to the inclusion of education as a quadratic term (Appendix Table A3). Standard errors are clustered by mother. *p<0.1 **p<0.05 ***p<0.01.

Table 3: Twinning and Stress *in Utero*

Dependent Variable:	(1)	(2)	(3)
Twins×100			
ETA Bomb casualities 1st trimester of pregnancy	0.002	-0.002	-0.002
	(0.006)	(0.006)	(0.004)
ETA Bomb casualities 2nd trimester of pregnancy	-0.010***	-0.010***	-0.010***
	(0.004)	(0.004)	(0.004)
ETA Bomb casualities 3rd trimester of pregnancy	-0.012*	-0.013*	-0.013**
	(0.007)	(0.008)	(0.006)
Observations	6,793,890	6,759,120	6,759,120
Year×month and province FE	Y	Y	Y
Socio-demographic controls		Y	Y
Provice-specific linear year-month trends			Y

Notes: Data consists of the Quintana-Domeque and Ródenas-Serrano (2017) sample of live births conceived between January 1980 and February 2003. Treatment is defined as number of ETA bomb casualties in the province of conception. Standard errors are clustered at the level of the province (50 provinces). *p<0.1; **p<0.05; ***p<0.01.

Table 4: Test of hypothesis that women who bear twins have better prior health

	(1)	(2)	(3)
	Birth 2	Birth 3	Birth 4

Panel A: Developing Country

Dependent Variable = Infant Mortality Rate×100

Treated	0.211	-1.283***	-1.722***
	(0.183)	(0.154)	(0.148)
Mean Value	9.983	10.443	11.159
Observations	542,186	422,498	312,350

Panel B: United States Birth Certificates

Dependent Variable = Miscarriage Rate×100

Treated	-0.727***	-0.238***	-0.063
	(0.050)	(0.053)	(0.067)
Mean Value	10.880	10.519	9.911
Observations	4,945,728	2,657,239	1,131,971

Notes: Dependent variables are constructed as the proportion of any prior births which have died in the first year of life (panel A), or resulted in miscarriage or fetal death (panel B). Regressions are run at the level of the mother taking averages over all prior births/pregnancies. In panel A, only children who have been entirely exposed to the risk of infant mortality are included (ie those over 1 year of age). Treated refers to giving birth to twins (rather than singletons) at the birth order indicated in the column header. A full description of these samples and the treatment variable is provided in section 2. Regressions include mother's age and race fixed effects. Standard errors are robust to heteroscedasticity. $*p<0.1$; $**p<0.05$; $***p<0.01$

Table 5: Fetal Deaths, Twinning, and Health Behaviours

Dependent Variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fetal Death × 1,000	Smokes	Drinks	No College	Anemic	N Cigs	N Drinks	Years Educ
Panel A: Uninteracted Twin – Non-Twin Difference							
Twin	9.979***	10.375***	10.397***	11.387***	9.971***	10.367***	10.397***
	(0.118)	(0.119)	(0.108)	(0.115)	(0.117)	(0.119)	(0.108)
Constant	5.344***	5.508***	5.172***	5.964***	5.337***	5.500***	5.172***
	(0.021)	(0.021)	(0.019)	(0.020)	(0.021)	(0.021)	(0.019)
Panel B: Health, Twin and Twin×Health Interaction							
Twin	9.907***	10.368***	8.991***	11.337***	9.939***	10.354***	19.630***
	(0.123)	(0.119)	(0.145)	(0.117)	(0.121)	(0.119)	(0.552)
Health (Dis)amenity	1.394***	4.924***	1.683***	0.608***	0.108***	0.602***	-0.242***

(0.066)	(0.260)	(0.038)	(0.131)	(0.005)	(0.038)	(0.007)
Twin × Health 1.154***	3.559**	3.573***	1.303**	0.061*	0.756***	-0.674***
(0.416)	(1.754)	(0.218)	(0.641)	(0.032)	(0.206)	(0.040)
Constant 5.195***	5.476***	4.268***	5.949***	5.214***	5.482***	8.277***
(0.022)	(0.021)	(0.028)	(0.020)	(0.021)	(0.021)	(0.088)
Observations 13,660,400	13,809,830	15,909,836	16,158,564	13,679,142	13,828,573	15,909,836

Each column in panel A represents a regression of whether a pregnancy ends in a fetal death (multiplied by 1,000) on whether the pregnancy is a twin pregnancy. Panel B augments the same regressions to include a health behaviour or health stock, and the interaction between being a twin pregnancy and the health variable. The health variable in each column is indicated in the column title. Regressions including controls for mother's age, child birth year and total fertility fixed effects are presented in Appendix Table A11. *p<0.10; **p<0.05; ***p<0.01.

Table 6: Can Selective Maternal Survival Explain Twinning Rates?

Dependent Variable: Twins×100	MMR Sample	<140cm or BMI <16	<145cm or BMI <16.5	<150cm or BMI <17	<155cm or BMI <17.5
Height	0.065***	0.063***	0.058***	0.051***	0.044***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
BMI	0.048***	0.045***	0.044***	0.041***	0.042***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Observations	844,552	848,552	848,492	848,468	848,507
R-Squared	0.024	0.024	0.023	0.023	0.022

Each column presents a regression of maternal characteristics on twinning following specification 1. Column 1 includes the full sample of women surveyed in countries where the DHS maternal mortality module is applied. In columns 2-5 we inflate the sample by the number of women who, according to our sister method calculations, would exist in the sample if it were not for the fact that they died in childbirth (this match assumes that a woman's health is a good proxy for her sister's health, and estimates will be less precise if this proxy is weak). However our measure of (sister) maternal mortality is very clearly decreasing in (respondent) height, see Figure A1. We then examine the coefficients of interest in the estimates of equation 1 under the extreme assumption that all less-healthy women who died were pregnant with twins, while all healthy women who died were not. We create a range of different binary distinctions of 'healthy vs less-healthy', using the available individual data on height and BMI, with cut-offs described in column headers. Heteroscedasticity robust standard errors are reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$